Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
Department of Psychology, University of Michigan
Ann Arbor, Michigan

DAVID A. GRANT, Associate Editor
University of Wisconsin

WILLIAM K. ESTES, Associate Editor
Indiana University

Consulting Editors

NORMAN H. ANDERSON, *University of California, Los Angeles*
E. JAMES ARCHER, *University of Wisconsin*
FRED ATTNEAVE, III, *University of Oregon*
CLETUS J. BURKE, *Indiana University*
PAUL M. FITTS, *University of Michigan*
FREDERICK C. FRICK, *Massachusetts Institute of Technology*
FRANK A. GELDARD, *University of Virginia*
JAMES J. GIBSON, *Cornell University*
CLARENCE H. GRAHAM, *Columbia University*
HAROLD W. HAKE, *University of Illinois*
LLOYD G. HUMPHREYS, *University of Illinois*
ARTHUR L. IRION, *Tulane University*
HOWARD H. KENDLER, *New York University*
DONALD B. LINDSLEY, *University of California, Los Angeles*
KENNETH MACCORQUODALE, *University of Minnesota*
QUINN MCNEMAR, *Stanford University*
L. STARLING REID, *University of Virginia*
KENNETH W. SPENCE, *State University of Iowa*
DELOS D. WICKENS, *Ohio State University*

CONTENTS

Experimental Alterations in the Effectiveness of Verbal Reinforcement Combinations: A. M. BUCHWALD 351

Guidance in Pursuit Tracking: D. H. HOLDING 362

Adaptive Processes Determining Proprioceptive Perception of Verticality: R. G. PEARSON AND G. T. HAUTY 367

The Effects of Syllable Familiarization on Rote Learning, Association Value, and Reminiscence: D. A. RILEY AND L. W. PHILLIPS 372

The Effects of Motivation and Probability of Reward on Two-Choice Learning: P. J. WOODS 380

Age Difference in the Acquisition and Extinction of the Conditioned Eyelid Response: H. W. BRAUN AND R. GEISELHART 386

Some Physiological Correlates of Hunger and Pain: B. T. ENGEL 389

The Concurrent Effects of Proactive and Retroactive Inhibition: R. J. SEIDEL 397

Variables Affecting Children's Performance in a Probability Learning Task: H. W. STEVENSON AND M. W. WEIR .. 403

Some Factors in Probability Matching: I. RUBINSTEIN 413

American Psychological Association

Vol. 57, No. 6 June 1959

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc., at Prince and Lemon Sts., Lancaster, Pa. The subscription rate per volume is $8.00, or $16.00 annually. Single copies are $1.50. Subscriptions, orders, address changes, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This Journal publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects, except when the latter are specifically oriented toward the extension of behavior theory. Experimental psychometric studies and studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology.

Normally, articles exceeding 20 printed pages cannot be accepted. Within this limit, the piecemeal experiment-by-experiment reporting of psychological research is discouraged, and the reporting of the data and conclusions of substantial segments of research efforts is encouraged. Specifically, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) *must* be presented in a single article. Special provision is made for the publication of Supplementary Reports and Replication Reports in articles of not more than 1.5 printed pages (see Editor's Note, this Journal, 1957, 53, 1–2).

Address all articles submitted for publication to the Editor, Arthur W. Melton, Department of Psychology, University of Michigan, Ann Arbor, Michigan. **Manuscripts *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc., as described in the *Publication Manual of the American Psychological Association.*** When in doubt about practices of the Journal, authors should examine a recent issue.

Articles are published in the order of their receipt, except in rare circumstances. Authors are supplied with 50 free offprints without covers. All of the cost of an author's alterations in galley proof is charged to the author. Priority in publication is given to articles whose authors assume the full cost of publication. In 1959 the publication cost is $30 per printed page. Authors of priority publications receive no free offprints.

To save printing costs, supplementary material to articles in this Journal are deposited with the American Documentation Institute. Copies may be obtained by ordering the Document Number given in the footnote reference to ADI. Orders should be sent to ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Orders *must* be accompanied by checks or money orders payable to the Chief, Photoduplication Service, Library of Congress.

ARTHUR C. HOFFMAN, Managing Editor
HELEN ORR, Promotion Manager
BEATRICE B. HAMLIN, Editorial Assistant

Second-class postage paid at Lancaster, Pa.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

© 1959 by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 57, NO. 6 JUNE, 1959

EXPERIMENTAL ALTERATIONS IN THE EFFECTIVENESS OF VERBAL REINFORCEMENT COMBINATIONS

ALEXANDER M. BUCHWALD

Indiana University

The experiments reported in this paper deal with response acquisition under three reinforcement combinations, RW, RN, and NW. The events R, W, and N consist of *E* saying "Right" (R), "Wrong" (W), or nothing (N) after *S*'s response. Use of a particular combination means that the first named event follows all correct responses and the second all incorrect responses. In several experiments (Buchwald, 1959; Buss, Braden, Orgel, & Buss, 1956; Buss & Buss, 1956) involving various response-selection tasks, and in which the instructions included no references to learning, correct responses, or the events R, W, and N, *S*s learning under RN have learned more slowly than other *S*s, and differences between *S*s learning under RW and those learning under NW have been minimal.

Explanations of these results have involved various hypotheses concerning the reinforcement values of the events R, W, and N. (The term *reinforcement value* refers to the effect of a particular event used as a reinforcer, i.e., to a parameter which determines the change in response probability when that event follows a response.) In one explanation (Buss, Braden, Orgel, & Buss, 1956; Buss & Buss, 1956) it was assumed that R is a very weak positive reinforcer, W a much stronger negative reinforcer, and that N has a zero reinforcement value. In a second explanation (Buchwald, 1959) it was assumed that when *S* is exposed to either RN or to NW, N acquires a reinforcement value opposite in direction to that of the event with which it is combined, and that the value acquired by N in the NW combination exceeds that which it acquires in the RN combination. These explanations account for acquisition data equally well, but the second one also led to correct predictions concerning the effects of an extinction procedure following acquisition under the various combinations (Buchwald, 1959).

EXPERIMENT I

Experiment I was designed to test several of the assumptions concerning reinforcement values. In contrast to earlier (single combination) experiments in which each *S* was exposed to only one reinforcement combination, Exp. I was a "multiple combina-

tion" experiment since each *S* was exposed to all three combinations in a single series of trials. This was accomplished by assigning each combination to a different pair of response alternatives, with all of these presented in a single series of trials.

This procedure makes possible a test of the different assumptions about the reinforcement value of N. Under the assumption that N maintains a zero reinforcement value the multiple combination procedure is essentially similar to a single combination procedure. Hence the results of this experiment should be similar to those of earlier experiments. However, under the assumption that N acquires a reinforcement value when *S* is exposed to RN or to NW, the present results should differ from those of earlier experiments.

Since *S*s are exposed to R, W, and N under conditions which make it difficult to discriminate the events assigned to various response alternatives, it is reasonable to assume that N will acquire a positive value for some *S*s, a negative value for other *S*s, and no reinforcement value for still others. These occurrences will tend to retard response acquisition under RN, under NW, and under both, respectively, as compared to a single combination situation. Further, the hypothesis that N acquires greater reinforcement value under exposure to NW than under exposure to RN implies that the present procedure will retard response acquisition under NW to a greater extent than it will affect acquisition under RN. Therefore, in the present experiment, acquisition under both RN and NW should be slower (relative to that under RW) than in earlier experiments, and (unless N acquires a positive reinforcement value for the great majority of *S*s) the previously found superiority of response acquisition under NW, compared to RN, should be lessened or should disappear.

To test these hypotheses and the assumptions which underlie them, the acquisition data obtained under the various combinations will be compared in the same respects as those in an earlier single combination study (Buchwald, 1959), and these results compared with those of single combination experiments.

Data obtained in the present experiment can also be used to test hypotheses concerning the relative reinforcement values of R and W. These hypotheses can be tested by estimating learning parameters for these events since under the present experimental conditions the relative values of these parameters will depend upon differences in the effectiveness of the reinforcing events. Since the present experimental arrangements are similar to those to which the mathematical model developed by Estes (1957b) applies, this model can be used to estimate the reinforcement parameters and to provide theoretical curves based on these estimates. The fit of these curves to the trial data can be used to test the hypotheses in question.

Method

Subjects.—The *S*s were 36 students enrolled in introductory psychology courses.[1]

Materials.—The materials, which have been more fully described in an earlier paper (Buchwald, 1959), consisted of three packs, each containing 216 3 × 5-in. cards. A pair of 40% association value (Hilgard, 1951) nonsense syllables placed one above the other, and a three-digit number, appeared on each card. Three different syllable pairs and nine different numbers were used. For any one pack three different numbers were assigned to each syllable pair, with the assignments for

[1] The author wishes to acknowledge the assistance of Solomon E. Feldman and of Georgia O. Windman, in running *S*s, and making computations, respectively.

different packs being made according to the principles of a latin square. All numbers appeared equally frequently in each pack; each member of a syllable pair appearing equally often in the upper and lower position with each of its assigned numbers. Each pack was separately randomized with the restrictions that each block of three cards contain all syllable pairs, and each block of 18 cards contain one of each nonidentical card in the pack. For each pack a different syllable pair was assigned to each reinforcement combination with each of the syllables in a pair being correct for half of the *Ss*.

Procedure.—The *Ss* were randomly assigned to card packs, correct-response syllables, and one of two *E*'s, subject to restrictions which balanced all of these in a factorial design.

Each *S* was run individually. The *S* was seated across a table from *E* and told that this was "a coding task" and that for each card he was to respond with the one of the two syllables shown which "goes with the number on the card." He was further informed that initially he might have to guess, but that "it will become clear to you as we go." A verbatim account of the instructions appeared in an earlier paper (Buchwald, 1959).

Each of the 216 cards was exposed to *S* one at a time. Following each response *E* took the action indicated below, recorded the response, and exposed the next card.

Following every response *E* said "Right" (R), or "Wrong" (W), or said nothing (N) according to the following plan: RW syllable pair, R for correct, W for incorrect responses; NW syllable pair, N for correct, W for incorrect responses; RN syllable pair, R for correct, N for incorrect responses. The indicated reinforcement event occurred following every appropriate response made by *S*. The various reinforcers were contingent only on the syllable chosen for the response, the numbers being merely irrelevant stimuli.

Results and Discussion

Inspection of the data indicated that differences due to card packs, correct-response syllables, and *E*'s, were small or nonexistent. Since these variables are of no interest in the present connection they were disregarded in the following analyses.

Comparison of data for the three combinations.—The number of incorrect responses given on each syllable pair was counted for each *S*. Summary statistics for this variable are shown in Table 1. As the statistics in Table 1 suggest, these data are badly skewed, particularly in the case of the RN pair.

TABLE 1

SUMMARY STATISTICS FOR NUMBER OF INCORRECT RESPONSES UNDER THE VARIOUS COMBINATIONS (72 TRIALS PER COMBINATION)

Reinforcement Combination	Mean	Mdn	*SD*	Range
RN	17.22	8.0	21.01	0–71
NW	10.36	7.0	10.52	1–52
RW	6.17	3.0	6.14	0–22

Over-all differences were tested using Friedman's analysis of variance by ranks (Siegel, 1956) and were found to be significant at the .01 level. The sign test applied to differences between the RW pair and each of the other pairs was also significant at the .01 level. However, the same test applied to the RN vs. NW data was not significant at the .05 level. To help guard against a Type II error the Wilcoxon test for paired replicates (Siegel, 1956), and the *t* test for paired measures, were also applied to the RN vs. NW data with the same lack of significance resulting. Since consideration of intra-individual data shows a median difference of three fewer incorrect responses for RW than for either RN or NW, and a median difference of zero for RN vs. NW, these results are not surprising. Note that the *mean* difference in number of incorrect responses between the RN and NW pairs does not stem from a majority of *Ss* making more incorrect responses on the RN pair. Seventeen of the *Ss* made more incorrect responses on the RN pair than they did on the NW pair, 16 *Ss* made fewer incorrect responses on the RN

TABLE 2

SUMMARY STATISTICS FOR LAST TRIAL ON WHICH AN INCORRECT RESPONSE OCCURRED UNDER THE VARIOUS COMBINATIONS

Reinforcement Combination	Mean	Mdn	*SD*	Range
RN	29.33	14.5	28.3	0–72
NW	32.16	31.5	21.0	2–72
RW	15.80	07.5	16.0	0–57

pair, and for 3 *Ss* there was no difference. In contrast, 25 of 36 *Ss* made more incorrect responses on the RN pair than on the RW pair, and the same number made more incorrect responses on the NW pair than on the RW pair.

As a second response measure the last trial on which an incorrect response occurred was tabulated for each syllable pair. Summary statistics for this variable are shown in Table 2. Again the data are markedly skewed.

Over-all differences for last incorrect trial are significant at the .01 level (Friedman test). The last incorrect trial occurred earlier on the RW pair than on either of the other pairs (both at the .01 level using the sign test), and earlier on the RN pair than on the NW pair (at the .05 level using the sign test). For individual *Ss* the median last incorrect trial occurred 9 trials earlier on the RN pair than on the NW pair, and 13.5 trials earlier on the RW pair than on the NW pair.

These data can be summarized as follows. The bulk of the *Ss* made fewest incorrect responses, and had their last incorrect trial earliest, on the RW pair. About half of the *Ss* made fewer incorrect responses on the RN pair than on the NW pair, while for the other *Ss* this pattern was reversed. The *Ss* who made more incorrect responses on the RN pair than they did on the NW pair tended to make many more incorrect responses on the RN pair, but for all *Ss* the last incorrect trial tended to occur earlier on the RN pair than on the NW pair.

When these results are compared with those of single combination studies several differences are apparent. First, the present finding that significantly more incorrect responses occur on the NW pair than on the RW pair is counter to the results of single combination experiments. In the three such experiments (Buchwald, 1959; Buss, Braden, Orgel, & Buss, 1956) in which error measures were used no such difference was found. Second, the present finding that the last incorrect trial occurred later on the NW pair than on the RN pair is unparalleled in single combination experiments. Previous studies using this measure, or comparable trials to criterion measures (Buchwald, 1959; Buss & Buss, 1956) have yielded the opposite result. In fact in five sets of data obtained from single combination studies there has been no finding of superior acquisition under RN, as compared to NW, on any response measure. Both of these differences are in the direction predicted on the basis of the assumption that N acquires reinforcement value during response acquisition under RN or under NW.

The first of the differences mentioned may, of course, be due to the inherently greater sensitivity of the multiple combination procedure where each *S* serves as his own control. But this argument concerning sensitivity should apply equally well to the difference between the RN and NW pairs. There is no reason why the present design should be more sensitive in one regard than the other.

On the basis of the foregoing it seems reasonable to conclude that the present results do deviate from single combination results in ways which are (grossly) predictable from the assumption that N acquires reinforcement value when *S* is

exposed to it as one of a combination of reinforcement events. Taken together with the earlier finding (Buchwald, 1959) that in a single combination experiment a typical extinction procedure (N for all responses) yields practically no extinction unless *S* has been exposed to RN during acquisition, these results provide strong support for the acquisition-of-reinforcement-value hypothesis.

Comparison of the data shown in Tables 1 and 2 suggests that there is some divergence of the two response measures used in this experiment. The data for individual *S*s reveal this more strikingly; many *S*s made more incorrect responses on the RN pair and had a later last incorrect trial on the NW pair. The reverse pattern did not occur. This suggests that there is some aspect of the procedure which leads to the occurrence of incorrect NW responses relatively late in the series of trials. Consideration of the procedure in the light of the assumption that the reinforcement value of N is dependent upon the reinforcement event which is an alternative to it suggested the following explanation. When *S* no longer makes incorrect responses on the NW and RW pairs, all responses are followed by either R or N. If *S* has not discriminated the different reinforcement contingencies associated with the several pairs, subsequent trials will be similar to those under a single combination RN procedure and N will acquire a negative reinforcement value. Since this negative reinforcer will follow all correct responses on the NW pair, the probability of these responses will decrease. If the series continues for enough trials *S* will produce an incorrect response on the NW pair. The occurrence of this isolated event, incorrect response—W, may lead to a correct discrimination of the reinforcement contingencies and therefore to a cessation of incorrect responses on the NW pair.

It follows from this explanation that the number of trials intervening between the next-to-last and the last incorrect responses should be greater for the NW pair than for the other pairs. The data are summarized in Table 3. The data

TABLE 3

SUMMARY STATISTICS FOR NUMBER OF TRIALS INTERVENING BETWEEN THE LAST TWO INCORRECT RESPONSES UNDER THE VARIOUS COMBINATIONS

Reinforcement Combination	Mean	Mdn	*SD*	Range	*N*[a]
RN	2.48	2.0	1.29	1–8	29
NW	10.06	7.0	8.67	1–29	35
RW	4.55	2.0	7.54	1–42	29

[a] Number of *S*s having at least two incorrect responses on that pair. Data from *S*s not having two incorrect responses were not used in calculating entries in this table.

differ among the three pairs at the 01. level of significance (Friedman test), with the NW differences being larger than those for RN and for RW, at the .001 and .02 levels, respectively (sign test). These results are not merely an artifact of the differing numbers of incorrect responses since only data from *S*s with at least two incorrect responses on each pair are included, and since similar results were obtained when *S*s for whom the next-to-last incorrect trial occurred very late in the trial series were excluded.

Estimation of parameters and curve fitting.—Estes and his associates have shown that a simple equation fits trial data obtained from experimental situations in which *S* is informed after each response as to whether or not it was correct (Estes, 1957a, 1957b; Estes & Straughan, 1954). The derivation of this equation involves the assumption that P_n (the probability of occurrence of the correct response on the nth trial) does not depend upon whether a correct, or an incorrect, response occurred on the $n-1$ trial. For the present RW pair this is tantamount to the assumption that R and W have equally large reinforcement values with opposite sign.

The applicable equation (Estes, 1957b) is

$$P_n = 1 - (1 - P_1)(1 - \theta)^{n-1} \quad [1]$$

where P_1 is the probability of correct responses prior to the first trial, and θ is the learning parameter. By estimating P_1 separately, it is possible to estimate θ

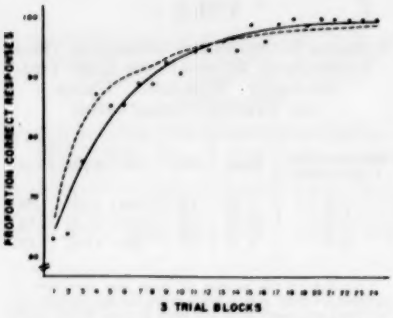

FIG. 1. Theoretical curves and empirical data for acquisition under RW. The solid line represents the curve with $\theta_R = \theta_W$; the dashed line represents the estimated curve with $\theta_R \neq \theta_W$; the solid circles represent the empirical data.

from the mean number of incorrect responses in a series of N trials.[2] Using $P_1 = .62$ (the proportion of initial responses on all pairs which were correct) yields as an estimate, $\theta = .0615$. These values were used in Equation 1 to produce the solid curve shown together with the empirical RW data in Fig. 1.

For purposes of comparison, separate estimates of the relative reinforcement parameters for R (θ_R) and for W (θ_W) were made. These estimates do not involve any assumption of equality of strength for R and W but they do require the assumption that the reinforcement value of N is zero in the present experiment. The estimate of θ_R was obtained from the mean number of incorrect responses over the 72 RN trials by use of equation 9.20 in Bush and Mosteller (1955, p. 198). The obtained value is .0278. Substituting this value in Bush and Mosteller's equation 9.20 (p. 198) yielded an estimate of $\theta_W = .2587$, based on RW data. This pair of values corresponds to her hypothesis that W is a much strongte reinforcer than is R. Since there is no known equation which is feasible to use to compute a theoretical RW curve from this pair of values, a theoretical curve was estimated by a Monte Carlo method designated as the use of "statrats" (Bush & Mosteller, 1955, pp. 129–131). The dashed curve shown in Fig. 1 represents the mean probability of correct responses obtained by running 10 statistical *S*s, each with $P_1 = .62$, $\theta_R = .0278$, and $\theta_W = .2587$, and behaving according to Estes' (1957b) model.[3]

Inspection of Fig. 1 suggests that the equal θ curve provides a somewhat better fit to the data than the alternative curve. Deviations of empirical data from the former curve are smaller on 20 of the 24 trial blocks than are the deviations from the second curve. Differences between actual and theoretical proportions range from .0005 to .0721 for the equal θ curve and from .0022 to .1261 for the latter curve, with the mean deviations being .0158 and .0250, respectively. Note in Fig. 1 that the fit of the unequal θ curve tends to be poorest where the two theoretical curves are most divergent. This is borne out by the results obtained by correlating differences between the two theoretical curves with deviations of empirical data from the unequal θ curve. The rank-difference correlation of .51 (significant at .02) indicates that the unequal θ curve tends to fit the data best when the unequal θ curve lies close to the equal θ curve. By contrast, the corresponding correlation for the equal θ curve is only .16 (nonsignificant). While these various indications of better fit are suggestive they do not seem sufficiently strong to warrant rejecting the hypothesis of unequal reinforcement values for R and W. Application of the runs test (Bush & Mosteller, 1955, p. 214) to direction of deviations of empirical data from theoretical curves (used as a test of goodness of fit) does not allow rejection

[2] By algebraic manipulation and summation of both sides of Equation 1

$$\sum_{n=1}^{n=N} (1 - P_n) = (1 - P_1)\,\frac{[1 - (1 - \theta)^N]}{\theta}$$

The left hand side is the expected number of incorrect responses in N trials.

[3] The responses consisted of entries from a table of random numbers. When the nth entry for an S was no larger than P_n, $P_{n+1} = (1 - \theta_R)P_n + \theta_R$. When the nth entry was larger than P_n, $P_{n+1} = (1 - \theta_W)P_n + \theta_W$.

of the null hypothesis at the .05 level for either curve.

Therefore, the results of this experiment (*a*) confirm the hypothesis that N acquires reinforcement value when *S* is exposed to this event in single combination procedures, (*b*) suggest that in a multiple combination procedure N may vary in value, perhaps in both directions, during the series of trials, and (*c*) are in accordance with, but at best weakly confirm, the hypothesis that R and W are equally potent reinforcers.

Experiment II

Experiment II was designed primarily to provide an additional test of the hypothesis that an originally neutral event, N, acquires reinforcement value when it is an alternative to some other event which already possesses a nonzero reinforcement value. Because this hypothesis is a relatively new one, it seems desirable to test it in a number of different ways.

There were four successive sets of trials. In Set I, "original learning," *S* was exposed to the task used in Exp. I under the RW combination until he reached a learning criterion. In Set II, "induction training," *S* was exposed to a revised version of this task under either NW, or RN. In Set III, *S* was tested for retention of the originally learned responses and given retraining if necessary. In Set IV, "reversal learning," *S* was given a series of trials on the original material, under multiple combination conditions, but with the responses which were correct in original learning now incorrect, and vice versa.

Sets I and III were intended to bring the probability of correct responses to zero (or near this figure) at the start of reversal learning. Induction training was used to introduce the experimental variables, reinforcement combination used in induction training, and number of trials of induction training. If the hypothesis that N acquires reinforcement value is correct, these variables will affect performance during reversal learning.

This design obviously requires that the task and materials used in induction training differ from those used in reversal learning. However, preliminary work suggested that if the tasks were made too dissimilar the reinforcement value acquired by N might not apply in reversal learning. To meet this difficulty, a revised version of the task used in Exp. I was devised for use in induction training.

Method

Subjects.—The *S*s were 84 introductory psychology students.[4] None of these had previously served in an experiment of this type.

Materials.—The materials used for all parts of this experiment except for induction training were identical with those used in Exp. I.

Three new packs of cards were constructed for use in induction training. The format of these cards was identical with that of the original cards. Six different syllable pairs were used in the new packs. For any *S* three of these pairs would consist of pairs of the three syllables which were correct in original learning, and three would consist of pairs of the three syllables which were incorrect in original learning. Each new pack was used in conjunction with only one of the original packs. These two packs were related in that a number assigned to a given syllable pair in the new pack had not appeared with either member of that pair in the original pack. Each of the new packs contained 48 cards arranged in random order subject to the restriction that each of the six syllable pairs appear once in each block of six cards.

Procedure.—The *S*s were randomly assigned to groups, card packs, and correct responses in original learning, subject to

[4] The author wishes to acknowledge the assistance of Muazzez O. Poffenbarger who ran these *S*s.

restrictions which jointly balanced card packs and correct responses over groups.

The instructions and basic procedure were identical with those used in Exp. I. The different parts of the experiment were run off, without any break, in the following order:

Original learning utilized the same procedure for all *Ss*. All correct responses were followed by R, all incorrect responses by W. Training was continued until *S* had given correct responses to all cards in an 18-card block. Only one *S* had to be discarded because he failed to reach this criterion within the previously set limit of 162 trials.

Induction training consisted of 12, 24, or 48 trials, with R for correct responses and N for incorrect responses (Groups R12, R24, R48), or with N for correct responses and W for incorrect responses (Groups W12, W24, W48). Group 0 received no induction training. During induction training each syllable was correct in one of the two syllable pairs in which it occurred, and incorrect in the other. For example, DIB was correct when paired with ZEW, and incorrect when paired with MOH.

Retention and retraining were carried out for all *Ss* except those in Group 0. The *S* responded until he had reached a criterion of three consecutive correct responses. Each *S* was run under the reinforcement combination to which he had been exposed during induction training.

Reversal learning was identical for all *Ss* except that the assignment of syllable pairs to reinforcement combinations was different for each card pack. During reversal learning, responses which had been incorrect during original learning were correct, and vice versa. The procedure was identical with that for Exp. I except that *S* was run for 48 trials on each of the three syllable pairs.

Results and Discussion

The number of incorrect responses made in original learning was subjected to analysis of variance to determine whether there were differences in the performance of the various groups prior to the introduction of the experimental treatments. All of the *F* ratios obtained from a simple analysis of variance on all seven groups, and from a 2×3 analysis on the six groups given induction training, were less than unity.

Summary statistics for number of incorrect responses during reversal learning are shown in Table 4. The data of primary interest are differences in the number of incorrect responses on the RN pair and on the NW pair. The statistical analysis was carried out on the square root of these differences since these variances were homogeneous while the variances of the untransformed differences were not. A 2×3 analysis of variance (Group 0 was omitted from this and subsequent analyses except where explicitly mentioned) applied to these data yielded significant *F* ratios for reinforcement combination used in induction training ($P < .01$), and for interaction ($P < .05$), but not for number of trials of induction training.

TABLE 4

NUMBER OF INCORRECT RESPONSES DURING REVERSAL LEARNING

Group	RN		NW		RW		RN-NW	
	Mean	Mdn	Mean	Mdn	Mean	Mdn	Mean	Mdn
R12	21.42	17.5	7.75	6.5	5.67	4.5	13.67	5.0
R24	6.83	3.0	10.50	11.0	7.33	3.5	−3.67	−3.0
R48	12.75	7.5	10.25	10.0	8.00	6.5	2.50	0.0
W12	21.83	17.0	7.50	5.5	6.62	5.5	14.33	5.0
W24	24.83	21.0	7.50	5.0	8.25	4.5	17.33	8.5
W48	35.83	39.0	7.17	5.0	6.62	5.0	28.67	21.5
0	6.92	1.5	5.58	2.5	3.12	2.0	1.33	−1.0

To ascertain whether these effects stemmed from experimentally produced differences in performance on the RN pair, or on the NW pair, or on both, analyses were also performed on the variables, NW-RW (number of incorrect NW responses minus number of incorrect RW responses), and RN-RW (number of incorrect RN responses minus number of incorrect RW responses). The RW data were used in this way in order to remove the effects of individual differences in speed of reversal from the NW, and RN data, and thus to decrease within-cell variances.

The NW-RW data were subjected to a 2×3 analysis of variance which yielded a significant F ratio only for reinforcement combinations used in induction training ($P < .05$). The RN-RW differences could not be subjected to analysis of variance since the variances were heterogeneous and it appeared unlikely that any transformation would lead to homogeneity. Therefore these data were tested by means of a nonparametric test (Wilson, 1956). The Wilson test yielded significant χ^2 values for reinforcement combinations ($P < .01$), and for interaction ($P < .001$). None of these results can be attributed to differences in performance on the RW pair since a 2×3 analysis of variance of the RW data yielded no F greater than unity.

These results can be summarized as follows. In general, the number of incorrect RN responses exceeded the number of incorrect NW responses, but these differences were greater for *Ss* given induction training under NW than for those given induction training under RN. The extent to which pairs of groups given identical numbers of induction training trials differed in this respect, was a positive function of number of induction training trials; the difference between groups R12 and W12 being virtually nonexistent. While all of these differences were largely due to the effect of induction training on performance on the RN pair, differences in performance on the NW pair also made some contribution.

Data consisting of the last trial on which an incorrect response occurred were also subjected to statistical analysis. The Wilson test was used for all tests on these data because of the extreme skewness of the data obtained from some of the groups. The data generally resemble the incorrect response data and since only minor differences in significance levels resulted, this material will not be reported.

In all of the above analyses data obtained from Group 0 have been omitted. It is apparent from Table 4 that *Ss* in Group 0 tended to make fewer incorrect responses on all syllable pairs than *Ss* in the other groups. Examination of data from individual *Ss* discloses that 6 of the 12 *Ss* in Group 0 made no more than two incorrect responses on any of the three syllable pairs, while only 4 of the 72 other *Ss* performed at this same level. These differences suggest that reversal learning immediately after original learning may differ qualitatively from reversal learning following a series of induction training trials. For this reason, no use is made of the data from Group 0.

The results of this experiment indicate that exposure to either RN or to NW affects the later course of learning under these reinforcement combinations. In general, the results are consistent with the hypothesis that a neutral event will acquire a reinforcement value opposite in sign to that of an event to which it is an alternative. However there are several aspects of the data which do not

seem to fit neatly within this interpretation. Two of these, the superiority of performance under NW in Groups R12 and W12, and differences in the extent to which induction training affected performance under NW and performance under RN, can be explained by considering the experimental procedure.

From the near identity of the data obtained from Groups R12 and W12, it can be concluded that 12 trials of induction training were insufficient to produce a nonzero reinforcement value for N (at least for most *S*s). When N has not acquired reinforcement value, its occurrence will not affect response probability (Buchwald, 1959). But since most responses made at the start of reversal learning will be incorrect, and since N follows incorrect RN responses while W follows incorrect NW responses, the probability of correct NW responses will be increased much more frequently than that of correct RN responses. Thus, a greater number of incorrect RN responses can be expected to occur over the entire series of trials.

These considerations also provide part of the explanation for the fact that induction training affected RN performance to a greater extent than it affected NW performance. Early in reversal learning the reinforcement value of N has only a minimal effect upon NW performance, but it is the chief determinant of RN performance. Further, when N has a positive reinforcement value the probability of correct RN responses can remain at zero throughout the series of trials since the occurrence of each incorrect response serves to prevent the occurrence of correct responses. But even if N should have a negative reinforcement value equal to that of W, the probability of correct NW responses can be expected to approach .5. (This is the most extreme case since it is reasonable to assume that W will have at least as large a negative value as N.) However, even in this case learning may proceed since the frequent alternation of N and W to members of the same syllable pair should serve to reduce, or even to reverse, the negative value of N. From these considerations it is apparent that the expected number of incorrect RN responses when N has a positive value is much greater than the expected number of incorrect NW responses when N is a negative reinforcer. But when N has reinforcement values which facilitate reversal learning under each of the combinations, the expected numbers of RN and NW incorrect responses are virtually identical. Hence, differences among groups in RN performance should be greater than differences in NW performance.

Finally, the discrepant results obtained from Group R48 remain to be considered. The discrepancies are twofold; namely, the failure to manifest as great an effect of induction training as was shown by Group R24, and the fact that the number of incorrect RN responses is not less than the number of NW responses. Any explanations of these results which might be offered at this stage would be unduly speculative. However, it should be noted that, despite the discrepancies, differences between R48 and all other groups are consonant with the hypothesis being investigated.

Summary

Two experiments designed to test various hypotheses about the reinforcing effects of three verbal events were performed. The verbal events, R, W, and N, consisted of *E* saying "Right" (R), "Wrong" (W), or nothing (N) after *S*s response. These were used in combinations, RW, RN, NW, in which the first named event followed all correct, and the second all incorrect responses. The instructions made no reference to these events.

In Exp. I, 36 *S*s were presented with 72 trials on each of three syllable pairs intermingled in a single series. Each of these pairs was assigned to one of the three reinforcement combinations. The results were as follows: (*a*) Fewest errors occurred on the RW pair while the other two pairs were not differentiated in terms of number of errors. (*b*) The last error occurred earliest on the RW pair, and next earliest on the RN pair. (*c*) The number of trials intervening between the last and next-to-last error was greatest for the NW pair.

In Exp. II, seven groups of 12 *Ss* first learned to select one of each pair of syllables to mastery under RW. Six groups were then given either 12, 24, or 48 trials on a similar task, with three of the groups (R12, R24, R48) having this induction training under RN and three (W12, W24, W48) under NW. The seventh group, Group 0, had no induction training. All *Ss* were then exposed to 48 trials on each syllable pair under conditions similar to those used in Exp. I except that the originally learned responses were now to be reversed. Reversal learning results revealed that induction training under RN led to fewer RN errors and more NW errors than did induction training under NW, and that the effect of induction training on number of RN errors and differences in number of RN and NW errors depended on the number of induction training trials. Group 0 made fewer errors on all pairs than did the other groups.

The results of these experiments seem to confirm the hypothesis that during exposure to RN, N becomes a negative reinforcer, while during exposure to NW, N becomes a positive reinforcer. There were also indications that the assumption that R and W are equally potent reinforcers may be slightly superior to the assumption that R is a much weaker reinforcer than W.

References

Buchwald, A. M. Extinction after acquisition under different verbal reinforcement combinations. *J. exp. Psychol.*, 1959, **57**, 43–48.

Bush, R. R., & Mosteller, F. *Stochastic models for learning.* New York: Wiley, 1955.

Buss, A. H., Braden, W., Orgel, A., & Buss, E. H. Acquisition and extinction with different verbal reinforcement combinations. *J. exp. Psychol.*, 1956, **52**, 288–295.

Buss, A. H., & Buss, E. H. The effect of verbal reinforcement combinations on conceptual learning. *J. exp. Psychol.*, 1956, **52**, 283–287.

Estes, W. K. Of models and men. *Amer. Psychologist*, 1957, **12**, 609–617. (a)

Estes, W. K. Theory of learning with constant, variable, or contingent probabilities of reinforcement. *Psychometrika*, 1957, **22**, 113–132. (b)

Estes, W. K., & Straughan, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.*, 1954, **47**, 225–234.

Hilgard, E. R. Methods and procedures in the study of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 517–567.

Siegel, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

Wilson, K. V. A distribution-free test of analysis of variance hypotheses. *Psychol. Bull.*, 1956, **53**, 96–101.

(Received July 2, 1958)

Journal of Experimental Psychology
Vol. 57, No. 6, 1959

GUIDANCE IN PURSUIT TRACKING[1]

D. H. HOLDING

Nuffield Department of Industrial Health, King's College, Newcastle upon Tyne, University of Durham, England

One of the prime difficulties in learning theory has always been how, in the first place, to get a given response to occur. The obvious solution, where the trainer literally guides *S* through the correct response, has received comparatively little experimental attention. One of the difficulties is that, unless *S* is set to attend to and identify himself with the externally manipulated response, the learning situation will be poorly structured for reinforcing the correct response. Thus, with animal *S*s, Thorndike (1898) failed to teach a string-pulling problem by guidance; and while Maier (1949) had some success in reorienting fixated rats by guidance, Haslerud (1951) later showed mechanical maze guidance to be less effective in changing existing habits than in establishing new ones.

The situation is clearer with human *S*s, where the appropriate set may be established by instructions. However, interest in guidance as a human training technique has tended to lapse since the early maze-learning studies of Carr and his associates (1930), Waters (1930) and Melcher (1934), although these results were promising. Von Wright (1957) has recently shown that one of the defects of these experiments is the lack of information about incorrect alternatives at choice points; where this information is given, it is possible to bring a task to full learning by guidance alone. A more serious defect is that maze learning is barely a skill, in that it requires less fine motor adjustment, and more rote learning of the task stimuli, than does a continuous adjustive skill like pursuit tracking. The purpose of the present series of experiments is to evaluate mechanical guidance as a training technique in a continuous sensorimotor skill, where it may be of use as a form of synthetic training.

METHOD

Apparatus.—The task was a form of that described by Graham, Baxter, and Browne (1951), providing pursuit tracking on a vertical scale by means of a rotary control, and modified to give automatic following when necessary.

The target pointer moved in a random sequence between 21 points spaced from 0 to 20 cm. up the vertical scale, reaching each point once during one cam revolution of .5 min. (Course A). An alternative Course B consisted of movement between the same positions in a different permutation. Speed of displacement between the positions was constant at about 9-cm./sec., with intervals between displacements of about .6 sec. The control was a 6-cm. rotary knob to the right of the display, with a ratio requiring movement round its circumference equal to the pointer displacement.

The error between target course and response was recorded by an integrator and counter, via the output shaft of a differential. To administer guidance, the differential was locked, and a train of compensating gears was inserted; the result was to give normal behavior to the display accompanied by accurate tracking movements of the control and its associated pointer.

Subjects.—Male students and young university staff were used. They were drawn

[1] This work was carried out on behalf of the Flying Personnel Research Committee of Great Britain, under R. C. Browne. The author wishes to thank D. J. Newell and Doris Weightman for statistical advice and computation, and N. E. Loveless for constructive discussion.

from various faculties, although most were medical.

General procedure.—Trials were in 2-min. tracking periods, separated by 3-min. rest pauses. All *Ss* beginning unguided tracking were allowed to handle the control before their first trial, to prevent artificially high error scores due to ignorance of the gear ratio and of the clockwise-for-up display-control relationship. The score was announced at the end of every scored trial, and took the form of integrated error-time units.

For guidance trials, the following pointer and the control knob were locked to the target course generator. The *Ss* were told that the machine would track itself perfectly, and warned that they would be expected to do it on their own later on. They were then asked to grip the control knob lightly, allowing the knob to turn their hand; they were also asked to watch the pointers during the run. In all conditions the run was started with the pointer in one of its middle positions.

Experiment I

The main object of Exp. I was to discover at what point, on the learning curve of *Ss* practicing normally on Course A, lay the scores of *Ss* who had received guidance. A control group (A) therefore performed three trials on Course A, while another group was tested for 2 min. on A after a 2-min. guidance run. These relatively short training trials were adopted in order to place the control scores on the steepest part of the learning curve, as it was essential to the design that these scores should differ significantly. Preliminary work had shown that the learning curve on this apparatus was comparatively flat.

Since the target cam repeated its rotation during a run, there was a theoretical possibility that rote learning of the target course would occur. A second control group (AB) was therefore transferred to Course B after one trial on A; while an equivalent guidance group (AB) was tested on B after guidance on A, to obviate the effects of specificity. A third normal-practice group (B) was given three trials on Course B, to provide evidence for any difference in difficulty between the courses which might complicate the comparison.

The 40 *Ss* were therefore allocated to five groups of eight, in order of appearance.

Results.—The scores in all experiments were transformed logarithmically. Course B showed a marked tendency towards greater difficulty than Course A, but analysis of variance of the six trials by the Control A and Control B groups failed to confirm the difference at the 5% level.

The scores for Guidance A fell at about the same point on the curve as those for Trial 2 of Control A, and an analysis of variance was carried out on Control A Trial 1 versus Guidance A and Control A Trial 2. This showed that there was no significant difference between the guidance test trial and the control Trial 2, while the control Trial 1 was significantly different ($P < .05$) from the other two; it was concluded that one guided trial was approximately equal to a normal voluntary 2-min. training period.

Reference to other data suggests that the amount learned during one 2-min. trial on this apparatus is quite large. The mean error for Trial 1 of Control A *Ss* works out at approximately 7.7 mm., improving to 6.5 mm. on the second trial.[2] An estimate of the total learning curve, obtained from four *Ss* who were able to return on other days to complete nine trials each, suggests that the final average error is unlikely to be less than about 4.6 mm. The benefit due to guidance (at 6.6 mm.) is thus over one-third of the estimated total learning.

TABLE 1

Guidance as Learning (Exp. I)

Source	*df*	*MS*	*F*	*P*
Control A 1 vs. A 2 and Guidance A	1	.0275	5.39	.05
Between A 2 and Guidance A	1	.0008	0.16	
Residual	21	.0051		
Total	23			

[2] The mean error at the display, in mm., is approximately 1/10th of the antilog of the scores used in the analyses.

TABLE 2

GUIDANCE IN A-B TRANSFER
(EXP. I)

Source	*df*	*MS*	*F*	*P*
Control B 1 vs. B 2, AB 2 and Guidance AB	1	.0161	2.27	>.05
Between B 2, AB 2, and Guidance AB	2	.0010	0.14	
Residual	28	.0071		
Total	31			

There was no significant difference between the Control AB Trial 1, and the Guidance AB and Control AB Trial 2; but this would not be expected if both were transferring to a more difficult course. Trial 1 of Control B was tested against the three scores falling at about the same point on the B curve—Controls B and AB Trial 2, and Guidance AB—with nonsignificant results. There was no significant difference *between* the last three, and it may be tentatively assumed that the transferred guidance was equivalent in training value to a normal training trial, although the evidence is not conclusive.

As a further check on the value of guidance, the test trials for Guidance A and AB were pooled for testing against Trial 2 scores of Controls A and AB. There was no significant difference ($F = .4$, 1 and 29 *df*). This was taken to support the efficacy of guidance.

EXPERIMENT II

As the evidence in Exp. I for transfer of guidance relied on nonsignificance, and therefore did not constitute proof, the transfer trials were repeated in the reverse direction. A further series of 20 *Ss* was divided into two equal groups. The control group practiced on Course B and subsequently transferred to A, while the experimental group was guided on B and subsequently tested on A.

Results.—The guided *Ss* transferred at least as well to Course A as did the control *Ss*. A difference between the guided test trial and the second control trial, in favor of the guided *Ss*, was not significant. There was a highly significant difference between the first control trial and the guidance test and second control trials.

The absolute amount transferred from B to A (Control A Trial 1 minus Control BA Trial 2) was greater than the amount transferred from A to B (Control B Trial 1 minus AB Trial 2), but in the absence of full learning curves for Course B it is not possible to state that difficult-to-easy transfer was superior to easy-to-difficult transfer.

EXPERIMENT III

This experiment attempts to assess the relative contributions of the kinesthetic and visual components of guidance, as a first stage in further analysis.

Eighteen *Ss* in one experimental group were blindfolded during the guidance run, receiving only kinesthetic information from the rotation of the knob. A further group of 18 *Ss* folded their arms while they watched the pointers tracking in unison; they were not prevented from watching the rotating control.

To confirm the findings of Exp. I, and to form a baseline for comparison with the kinesthetic and visual groups, control and full guidance groups of 10 *Ss* each were again included.

All runs were on Course A.

Results.—Analysis of variance showed that it was unobjectionable to combine the scores of control and full guidance *Ss* in Exp. III with those

TABLE 3

GUIDANCE IN B-A TRANSFER
(EXP. II)

Source	*df*	*MS*	*F*	*P*
Control BA 1 vs. BA 2 and Guidance BA	1	.0851	16.26	.001
Between BA 2 and Guidance BA	1	.0089	1.71	
Residual	27	.0052		
Total	29			

TABLE 4

VISUAL AND KINESTHETIC GUIDANCE (EXP. I & III)

Source	*df*	*MS*	*F*	*P*
Kinesthetic effect	1	.0077	1.74	
Visual effect	1	.0334	7.56	.01
K-V Interaction	1	.0010	0.23	
Residual	68	.0045		
Total	71			

of the equivalent groups in Exp. I. With the combined groups of 18 *S*s in each category, the difference between the first control trial and the guidance test trial emerged as significant ($P < .01$). The difference between the second control trial and the guidance test trial remained not significant.

Using the combined data, the improvement over the first control trial due to full guidance was approximately 10 mm., while that due to visual guidance was 6 mm., and kinesthetic guidance 2 mm. of mean error. The data were treated factorially; the visual effect was significant ($P < .01$), while neither the kinesthetic effect, nor the kinesthetic-visual interaction was significant.

DISCUSSION

In this kind of tracking an amount of guidance has an effect roughly equal to a similar amount of normal practice, early in training, even though normal learning was encouraged by pretest familiarization and knowledge of score. The work has been restricted to early training; it is extremely difficult to prove the effect of guidance after greater amounts of learning, since the chances that points further down the learning curve will differ significantly become successively reduced.

It is far from clear that the effects of guidance are limited to the early stages of learning; Carr's (1930) conclusions on the relative decrease in effectiveness of larger amounts of guidance are not confirmed by Waters (1930). Obviously, as in any other case of transfer, the effectiveness of guidance will vary with its degree of simulation of the final task, and this will affect its efficiency more in the later stages of training. In any synthetic training device, overlearning on the trainer tends to impede transfer to the real task. In any case, where a decrease is relative to a large initial advantage, the early benefit will be of use in many training situations, where the early, fumbling, stages of learning are to be avoided. Where an operation may be performed in several different ways, guidance has the practical advantage of obviating the fixation of inefficient techniques; again, where a task to be learned conflicts with the existing repertoire of *S*, as in learning an adverse display-control relationship, guidance may have the advantage.

There is no reason to think that guidance will not be effective for a wide range of pursuit tasks. The amount of control movement which is usually necessary is well above the level for accurate discrimination of passive movement, which tends to be below 1° of movement under most conditions (Cleghorn & Darcus, 1952). On the other hand, it may be predicted that the benefit on a compensatory task will be slight, as compensatory guidance would mimic the kinesthetic condition of Exp. III.

The failure of the kinesthetic effect to reach significance was probably due to the fact that the guidance procedure was *S*'s first experience of the tracking task, and it seems likely that it would have benefited from insertion after a preliminary tracking trial. Poulton's (1957) comparison of obscuring the input or the response pointer shows clearly that knowledge of task requirements is prior to knowledge of the means; while Lincoln's (1956) moderate success in teaching rates of movement, with interpolated kinesthetic examples, is an indication that the contribution of kinesthetic guidance need not be nil.

The success of the visual technique is in accordance with both Melcher's (1934) findings and those of Twitmyer (1931).

Since, in the full guidance case, learning by passive movement does take place, it is no longer possible to hold that motor learning cannot occur without knowledge of results. The model of *S* as a servo, attaining motor adjustments through successive mismatches, does not fit the guidance procedure. We must obviously, with Annett and Kay (1957), distinguish between knowledge of results, as the feedback component of the servo model of skill, and extraneous knowledge of score. To the extent that knowledge of results is equated with error feedback, two differences are obvious in the case of guidance: there is no error to feed back, since the automatic tracking is perfect; and there is no feedback, since the kinesthetic stimuli of guidance are not in fact fed back as the result of any efferent quantities.

The theory that knowledge of results is indispensable conceals the distinction between knowledge of results and knowledge of the correct response. What seems to be necessary for the acquisition of motor skill is the occurrence of the correct response; and knowledge of results is merely an incidental, though common, means of achieving the correct response.

Summary

Three experiments on pursuit tracking were conducted with 116 *Ss*. The apparatus was modified so that guidance could be administered mechanically.

It was found that guidance at an early stage produced learning to about the same extent as did normal practice. This was the case when a single target course was used; and when *Ss* transferred learning on a first target course to tracking a second. Transfer was carried out in both directions between the tasks. Full guidance was more effective than visual guidance; and visual guidance was superior to kinesthetic guidance, which gave no significant benefit in this instance.

These results suggest a reinterpretation of the place of knowledge of results in the theory of motor skill, and a new evaluation of guidance as a training technique.

References

Annett, J., & Kay, H. Knowledge of results and 'skilled performance.' *Occup. Psychol.*, 1957, **31**, 69–79.

Carr, H. A. Teaching and learning. *J. genet. Psychol.*, 1930, **37**, 1189–218.

Cleghorn, T. E., & Darcus, H. D. The sensibility to passive movement of the elbow joint. *Quart. J. exp. Psychol.*, 1952, **4**, 66–77.

Graham, N. E., Baxter, I. G., & Browne, R. C. Manual tracking in response to the display of horizontal, vertical and circular scales. *Brit. J. Psychol.*, 1951, **42**, 155–163.

Haslerud, G. M. Anticipative transfer of mechanically guided turns. *J. exp. Psychol.*, 1953, **45**, 431–436.

Lincoln, R. S. Learning and retaining a movement with the aid of kinesthetic and verbal cues. *J. exp. Psychol.*, 1956, **51**, 199–204.

Maier, N. R. F. *Frustration.* New York: McGraw-Hill, 1949.

Melcher, R. T. Children's motor learning with and without vision. *Child Develpm.*, 1934, **5**, 315–350.

Paulton, E. C. On the stimulus and response in pursuit tracking. *J. exp. Psychol.*, 1957, **53**, 189–194.

Thorndike, E. L. Animal intelligence. *Psychol. Rev. Monogr. Suppl.*, 1898, **2**, No. 8.

Twitmyer, E. M. Visual guidance in motor learning. *Amer. J. Psychol.*, 1931, **43**, 165–187.

von Wright, J. M. A note on the role of guidance in learning. *Brit. J. Psychol.*, 1957, **48**, 133–137.

Waters, R. H. The influence of large amounts of manual guidance upon human maze learning. *J. gen. Psychol.*, 1930, **4**, 213–227.

(Received May 26, 1958)

Journal of Experimental Psychology
Vol. 57, No. 6, 1959

ADAPTIVE PROCESSES DETERMINING PROPRIOCEPTIVE PERCEPTION OF VERTICALITY

RICHARD G. PEARSON

Carnegie Institute of Technology

AND GEORGE T. HAUTY

USAF School of Aviation Medicine

Perception of postural verticality as a function of learning has received little primary attention (Fleishman, 1953; Passey & Guedry, 1949) until recently when Solley (1956) presented positive evidence of the effect of such learning. This was in the form of a clearly defined negatively accelerated curve and was obtained by plotting the magnitude of perceptual error for 30 successive trials of lateral tilt.

Of particular interest to the present authors is the fact that following each trial judgment *S*s were returned from their judged vertical position to the true vertical and told they were then "upright." Clearly, such procedure would seem to provide some degree of knowledge of results (KR) and, equally clear, is the possibility that this might have been largely responsible for the rate and degree of learning achieved.

The present study therefore was designed to determine the contribution made by KR presented in this proprioceptive manner, the comparative effectiveness of KR presented by visual means, and the nature of the interaction of these presumed effects with alternation of the direction of lateral tilt.

METHOD

Apparatus.—The lateral-tilt chair and its essential components have been described elsewhere (Pearson, 1957). A shutter mechanism, an integral part of the chair, was used only in conjunction with the visual KR device shown in Fig. 1. This device, wall-mounted opposite the shutter, consisted of a transparent disk pinned at its center to a fixed black board. On the latter was painted an error scale, true vertical, and horizon indicators. The perceived vertical indicator was painted on the plastic disk which, when rotated by *E*, clearly indicated the magnitude of departure from true vertical. The paint, a luminous substance, was energized under red light. The *S*'s control panel, modified for the present study, contained adjustable leather sleeve cuffs to support *S*'s forearms at tilt positions and a three-position switch providing *S* with directional control of the chair which, of course, was also under the independent control of *E*.

The test room was finished with a flat black paint and was completely light proof. Both red and white light were available to *E* at his discretion. Temperature, humidity, and ventilation were adequately controlled within the room.

Design.—The experimental conditions were arranged in conventional factorial design involving two variables, KR and alternation of tilt. The former consisted of the following subclasses: KR_1—no KR (control) in which

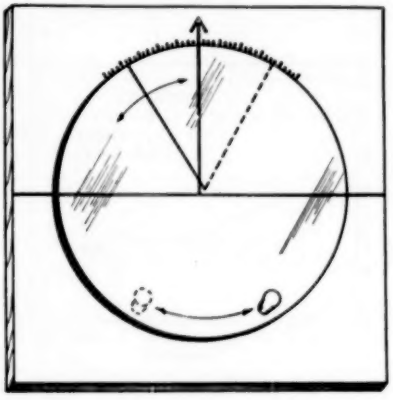

FIG. 1. Visual KR apparatus.

S was moved directly to tilt following his trial judgment; KR_2—postural KR in which S was moved to the true vertical following each judgment; KR_3—visual KR in which S was shown a visual representation of his position following each judgment. The second variable consisted of A_1—24 trials at 30° left tilt; A_2—24 trials alternately 30° left and 30° right tilt. The *Ss* were assigned in equal numbers and randomly to each of the six cells.

Procedure.—The S was secured in the tilting chair following adjustment of the seat to his sitting height, given detailed instructions, and told he would receive 24 trials. No mention was made of the actual degree of tilt. If S had no questions, testing then commenced. The E moved S to a position of tilt, then, following a delay period of 30-sec., transferred control of the chair to S who made his judgment. When satisfied with his position, S responded "OK" or "All right." The E then recorded the direction and amount of error. Delay at the *perceived* vertical was 20 sec. For the visual KR *Ss*, E, after setting the KR device, opened the shutter for 5 sec. during the latter half of the delay period. In the case of the postural KR, S's delay at the *true* vertical was 10 sec. Thereafter, E proceeded in accordance with the KR condition to which S was assigned. All 24 trials were completed in one session.

Subjects.—The *Ss* were obtained from a population of volunteer, male, basic Air Force trainees. Four were tested per day until a sample of 96 was attained.

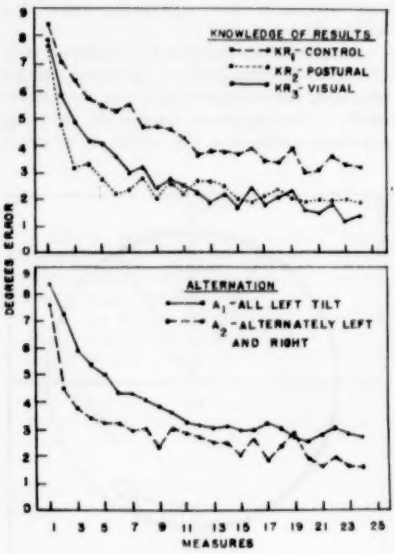

FIG. 2. Main effects of KR and Alternation upon perception of the postural vertical.

RESULTS

Consideration was first given to the absolute magnitude of error in proprioceptive perception of the vertical. Figure 2 plots the average errors for the KR and Alternation main effects. One is immediately impressed with the reduction in error that occurs for all conditions as testing progresses. Considering first the upper half of Fig. 2 one sees the beneficial effects of KR upon performance. Postural KR initially has a greater positive effect than visual KR, but eventually (about Trial 8) the visual subgroup attains the same performance level as the postural subgroup. Turning now to the lower half of Fig. 2, it is to be noted that the alternation (A_2) subgroup has a marked early reduction in error as compared with the control (A_1) subgroup, and, in general, maintains this superior performance throughout testing. The above findings are supported by the statistical results, shown in Table 1, based upon a repeated measurements analysis made withing the split-plot design of KR (3) × Alternation (2) × Measures (24). It is to be noted that all main effects and their first-order interactions are statistically significant.

In Fig. 3, the principle course of these significant interactions is clearly resolved. Under the A_1 conditions, the no-KR group evidences a near linear reduction in error whereas the KR groups produce the curves of a quadratic function. Under the A_2 condition, on the other hand, the no-KR group performs quite differ-

TABLE 1

ANALYSIS OF VARIANCE OF ABSOLUTE ERRORS

Source	*df*	*MS*	*F*
KR	2	781.73	10.49***
Alternation (A)	1	634.94	8.52**
KR × A	2	322.92	4.33*
Error (a)	90	74.52	
Measures (M)	23	177.99	39.12***
M × KR	46	6.56	1.44*
M × A	23	10.87	2.39**
M × KR × A	46	4.18	0.92
Error (b)	2,070	4.55	
Total	2,303		

* $P < .05$.
** $P < .01$.
*** $P < .001$.

ently. Its curve is scarcely distinguishable from the quadratic curves of the KR groups.

It should also be mentioned that two additional repeated measurements analyses were performed upon the average errors. In one, only the data from the odd-numbered trials were analyzed; here there was a comparison of measures made only at the left-tilt position and equated in time. In the second analysis the data from the first 12 trials for the A_1 condition were compared with that from the odd-numbered trials for the A_2 condition; here we equated for the number of successive measures made at left tilt. Suffice it to say that the results of these analyses were identical with the previously described one, namely, that all main effects and their first-order interactions were statistically significant.

Finally, the authors gave consideration to the directional aspects of error in perception of the vertical. The results of plotting the constant errors for the odd-numbered trials were almost identical, in terms of slope and of the relative position of the curves, with those from the plotting of the absolute errors. Also, the constant errors for the even-numbered trials, right-tilt data only, were plotted. Here, in the case of both the Alternation main effects and the interaction effects (KR_1A_2, KR_2A_2, and KR_3A_2), the performance curves were essentially of the same slope as their "left-tilt" counterparts. Analyses similar to those performed on the average error data were performed on the constant error data. Without going into detail it can be said that the statistical results supported the graphic findings, and, in general, confirmed the results obtained with average errors.

DISCUSSION

It has been shown that *Ss*, with repeated testing, will demonstrate a significant reduction of error in perception of the vertical. This is even true for

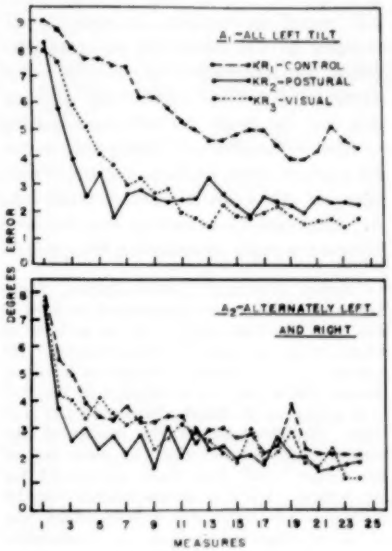

FIG. 3. Interaction effects of KR and Alternation upon perception of the postural vertical.

Ss who are tilted repeatedly in the same direction and who receive no knowledge of their performance. To the authors this evidence suggests that the early effect of sensory adaptation to tilt is to a degree overcome, as testing progresses, by proprioceptive learning.[1]

It was further found that when *Ss* are tilted repeatedly in one direction, both postural and visual KR served to bring about improved performance as testing progressed. Inasmuch as returning *S* to the true vertical following each trial judgment does have a positive effect upon performance, it is suggested that careful consideration be given to the inclusion of postural KR in the design of similar studies.

Most noteworthy were the findings with respect to the effects of "alternation" upon performance. It was shown that the performance of a group of *Ss* tilted alternately left and right was significantly superior to that of a group tilted repeatedly to the left. Furthermore, the effect of alternation was such that neither visual nor proprioceptive KR could contribute anything additionally to the reduction in error provided by alternation alone. In other words, under these conditions, alternation was the main variable contributing to the reduction of perceptual error. To explain these findings one must note that on all even-numbered trials the A_2 (alternation) subgroup has received different sensory stimulation than the A_1 subgroup. In Fig. 2 note especially the magnitude of the initial divergency between the subgroups at Trial 2 where, it is to be remembered, the only difference between them here is the direction of tilt. It would seem that in the case of the A_2 subgroup some of the sensory adaptation occurring at Trial 1 has been carried over to Trial 2, and thus has partially mitigated, or competed with, the sensory adaptation occurring at right tilt. Similarly it might be argued that at Trial 3 adaptation to left tilt for the A_2 subgroup has been partially mitigated by the previous adaptation to right tilt. Yet, another explanation seems to be equally plausible and is inherent in the fact that the A_2 subgroup on even-numbered trials receives "information" over and above that received by the A_1 subgroup; in other words, the A_2 subgroup, in *passing through the vertical*, enjoys a unique experience in that it learns what it is like to be on the other side of true vertical. This additional information may be expected to modify *S*'s subsequent judgments. Of course the possibility should not be overlooked that both experiential factors and competition between sensory events may jointly contribute to *S*'s increased efficiency in perception of verticality as testing progresses. In any case the above findings suggest that the practice of moving *S* to the opposite quadrant before moving him to the trial position (Mann, Berthelot-Berry, & Dauterive, 1949; Solley, 1956) cannot be considered just a precaution against the possibility that *S* would count himself back to the true vertical, but must be considered a significant part of a study's methodology which therefore can play an important role in determining performance.

A final point carries another implication for studies of spatial orientation. Solley's conclusion that ". . . use of so-called 'random orders' of presentation of conditions in kindred perceptual experiments should be looked upon with suspicion" (Solley, 1956, p. 332) is well made. In the light of our findings with regard to proprioceptive learning it is further apparent that the practice of

[1] Solley (Personal communication, January, 1959) may have anticipated us here in the case of what appears to be a kind of spontaneous perceptual improvement. He reports ". . . a striking amount of learning, though there was no knowledge of results" in a group of *Ss* tilted alternately left and right. On the basis of the results of our present study we feel that the mere fact of alternation itself may have accounted for his results. Until a determination can be made as to which variable, sensory adaptation or proprioceptive learning, is responsible for the improvement observed in our study, we must regard the question of whether perceptual learning can occur without specific knowledge of results as merely an hypothesis.

averaging the results of several consecutive trials made under certain conditions should also be looked upon with suspicion. For example, under conditions where *Ss* are tested in both right and left quadrants, it may well be that with repeated testing the effects of such variables as delay at tilt and speed of chair movement would cease to have a significant effect upon performance. It is noteworthy that the performance curves for the Alternation variable of the present study have approximately the same slope as those from Solley's experiment although there are several basic differences in design methodology between the two studies.

Summary

Ninety-six *Ss* were tested on a lateral-tilt chair under darkroom conditions. The task consisted of 24 trials and required *S* to return himself from a position of 30° tilt to the perceived vertical. It was found that: (*a*) *Ss* with repeated testing demonstrated a reduction of error in perception of the vertical, probably as a result of proprioceptive learning; (*b*) return to the true vertical following each trial judgment (postural knowledge of results) and visual representation of error (visual knowledge of results) both had a significant positive effect upon reduction of error; (*c*) *Ss* tilted *alternately* left and right performed significantly superior to *Ss* tilted *repeatedly* to the left. In the case of the alternation *Ss* it was further found that neither visual nor proprioceptive knowledge of results were contributing anything additionally to reduction of perceptual error than that provided by alternation alone. An attempt was made to explain this finding in terms of experiential factors and competition between sensory events resulting, in the case of the alternation group, from the mere process of passing through the true vertical.

Finally, implications of the findings for the design of spatial orientation studies were discussed.

References

Fleishman, E. A. Perception of body position in the absence of visual cues. *J. exp. Psychol.*, 1953, **46**, 261–270.

Mann, C. W., Berthelot-Berry, N. H., & Dauterive, H. J., Jr. The perception of the vertical: I. Visual and non-labyrinthine cues. *J. exp. Psychol.*, 1949, **39**, 538–547.

Passey, G. E., & Guedry, F. E. The perception of the vertical: II. Adaptation in four planes. *J. exp. Psychol.*, 1949, **39**, 700–707.

Pearson, R. G. The effects of motion-sickness preventives upon orientation in space. *USAF Sch. Aviat. Med., Rep. No. 58-7.* November, 1957.

Solley, C. M. Reduction of error with practice in perception of the postural vertical. *J. exp. Psychol.*, 1956, **52**, 329–333.

(Received June 3, 1958)

Journal of Experimental Psychology
Vol. 57, No. 6, 1959

THE EFFECTS OF SYLLABLE FAMILIARIZATION ON ROTE LEARNING, ASSOCIATION VALUE, AND REMINISCENCE [1]

DONALD A. RILEY AND LAURA W. PHILLIPS

University of California

In the two experiments reported in this paper, the effects of syllable familiarization on three separate phenomena were investigated. These were: (*a*) syllable association value, (*b*) speed of learning and the nature of error production in a serial rote learning task, and (*c*) reminiscence.

The first two phenomena were examined to test hypotheses concerning the reason familiarization reduces trials to learn. In their paper, Hovland and Kurtz (1952) suggested three different ways by which the effects of familiarization could influence learning speed. These three ways are: (*a*) by increased differentiation among the items within a list (cf. Gibson, 1940); (*b*) by increased differentiation between familiarized and unfamiliarized items; (*c*) by a reduction in the pool of alternative responses available to be anticipated. A proposal similar to (*c*) has also been made by Noble (1955). Each of these hypotheses has implications for the nature of errors made in learning familiarized items.

An additional hypothesis not considered by Hovland and Kurtz is that familiarization produces its effect through increasing the availability of meaningful mediators. The mediators in turn would facilitate the establishment of linkages between the syllables in a series. Specifically, this hypothesis would assert that familiarization would speed learning by increasing the association value of familiarized syllables. Although the effect of association frequency (Noble's *m*) on learning has been well-documented by Noble (1952a, 1952b) and by Mandler and Huttenlocher (1956), surprisingly little is known concerning the experimental variables that can produce changes in association value. It is likely that if words are paired with syllables in a familiarization procedure, the association value of the syllables will increase. It is less obvious but possible that sheer familiarization might also influence association value. By increasing the strength of the associative connections between the letters in a syllable, unitary responses to these letter combinations may develop. These responses may in turn elicit meaningful associates that act as mediators. Conversely, it is also possible that one of the ways by which a subject recalls the letters that form syllables is through an active search for meaningful associates that aid in the recall of all the letters in a syllable. Either process could affect association value. Regardless of the mechanism, however, increase in association value is clearly a possible result of familiarization. That changes in association value of syllables may occur without explicit pairing of syllables with meaningful items was demonstrated by Hull (1933). He showed that a few trials of rote learning increases the number of nonsense syllables that elicit meaningful associates. Therefore, in view of the known relation-

[1] This study was supported by a grant from the National Science Foundation.

ships between association value and learning speed on one hand, and between familiarization and learning speed on the other, the relationship between familiarization and association value is of systematic interest, both as a problem in its own right and because of its bearing on the familiarization-learning problem.

The other problem that was investigated was the effect of the familiarization procedure on performance following a rest interval at the point of approximately 50% learning. The procedure followed was a combination of those used by Ward (1937) and by Hovland (1938). Meaningfulness (cf. Melton & Stone, 1942) and more directly, interlist and intralist interference as investigated by Underwood and Richardson (1955) and by Riley (1957), have occupied central positions in the analysis of distribution of practice and reminiscence effects. These studies suggest that the extent to which a rest interval benefits further performance depends on the degree of interference prior to the rest. If familiarization influences error occurrence prior to a rest interval, it should also exert a corresponding effect on postrest performance.

METHOD

Materials

In both Exp. I and Exp. II, the experimental materials consisted of six practice lists and four critical lists. The practice lists were used to train *Ss* in the technique of serial learning. The critical lists were used to test the hypotheses discussed in the introduction.

The practice lists were 12-item serial lists constructed from Glaze's (1928) syllables. The lists averaged between 25% and 30% association value. The syllables in each list were selected so as to minimize intralist similarity. Within each list, all syllables had different first letters; all had different last letters. No vowel appeared more than three times in any list. No syllable appeared in more than one list. The same practice lists were used in both experiments.

In Exp. I each of the four critical lists was a 12-item serial list. The syllables were from Glaze's lists. These lists averaged between 19% and 22% association value. The rules of list construction were the same as for the practice sets. There was, however, one additional restriction: no syllable had the same first two letters or the same last two letters as any other syllable in any of the four critical lists. Between 16 and 18 consonants were used in the construction of each list. No critical list syllables were taken from the practice lists.

In Exp. II the critical lists were constructed according to the same rules as the critical lists of Exp. I. These new lists differed from those of Exp. I in that each list had 15 syllables rather than 12 syllables. This increase resulted in more repetition of consonants, but such repetition was held to a minimum. Between 17 and 20 consonants were used in the construction of each list.

Experiment I

For each of 20 paid male student *Ss*, the experiment extended over seven daily sessions of 1 hr. each. During Days 1–4, *Ss* memorized the six serial practice lists. In addition, *Ss* were given special familiarization training on the syllables of two of the four critical lists. On Day 5, *Ss* were given an association test on the items of all four critical lists. On Days 6 and 7, *Ss* learned the four critical lists.

Practice sessions.—Each *S* learned the six practice lists by the serial anticipation method. One list was learned on each of the first two days; two lists were learned on Day 3, and two on Day 4. The *Ss* were required to spell out their anticipations and to correct erroneous responses on seeing the correct response. The items appeared at a 2-sec. rate on a memory drum, with 6-sec. between trials. Each *S* learned three of the lists with a 2-min. rest interval following the trial on which seven syllables were first correctly anticipated. During the interval, *Ss* named colors that appeared at a 2-sec. rate on the memory drum. The other three lists were learned without the rest interval. Four orders of counterbalancing the learning of rest (R) lists and nonrest (NR) lists were used, of which R, NR, NR, R, NR, R is an example. Five *Ss* followed each schedule.

Syllable familiarization.—During the four practice days, each *S* learned to recognize the 24 syllables that comprised two of the four

TABLE 1

A SAMPLE GRECO-LATIN SQUARE USED IN THE ASSIGNMENT OF *Ss* TO CONDITIONS

S	ORDER			
	Day 6		Day 7	
	1	2	3	4
1	List B NR-U	List C R-U	List A R-F	List D NR-F
2	List C R-F	List D NR-U	List B NR-F	List A R-U
3	List A NR-F	List B R-F	List C R-U	List D NR-U
4	List D R-U	List A NR-F	List C NR-U	List B R-F

critical lists. The choice of the particular lists on which each *S* received this familiarization training was determined by *S*'s random assignment to one of four rows in one of five greco-latin squares.

Familiarization training consisted of showing *S* a series of 3 × 5 in. cards, each one of which had on it either the first two or the last two letters of a nonsense syllable. For each syllable there were two cards, one of each variety. As each card was presented, *S* was required to spell out all three letters of the syllable from seeing only two. After approximately 3 sec. the card was turned over and *S* was shown all three letters. At the end of a trial, those cards which had been correctly recognized were dropped from the pack, and the remaining cards were shown in the same manner on the next trial. This procedure was continued until all the cards had been correctly recognized. Then all 48 cards were shown again in the same way. This procedure was continued until all 48 syllables had been correctly completed on two successive errorless trials. The cards were shuffled between each exposure of the remaining items in the pack. All *Ss* attained the level of proficiency just described prior to Day 5. At the beginning of Day 5, each *S* was again required to recognize all 48 cards in one trial, the same procedure being followed.

Association test.—Following the completion of familiarization, *Ss* were shown, in random order, a series of 48 cards, each with all three letters of a syllable on it. Twenty-four of the cards had the familiarized items, the others had the unfamiliar syllables from the other two critical lists. As each card was presented, *S* was required to respond as rapidly as possible with as many words as the syllable reminded him of, for a period of 30 sec. He was instructed to give only words suggested by the syllable and not to give chains of associations (Noble, 1952a). The *E* recorded both the time to the first response and the responses given.

Critical lists.—On Days 6 and 7, *S* learned the four critical lists. The procedure was the same as for the practice lists. One of the lists containing familiarized syllables and one of the lists containing nonfamiliarized syllables were learned with a 2-min. rest following the trial on which seven syllables were first correctly anticipated. The other "familiar" list and the other "unfamiliar" list were learned with the regular 6-sec. break at that point. These four conditions are hereafter designated NR-F (nonrest, familiar), R-F (rest, familiar), NR-U (nonrest, unfamiliar), and R-U (rest, unfamiliar). Two lists were learned each day. The order in which each *S* learned the four critical lists was determined by the same set of greco-latin squares that determined the pairings of *Ss* and lists. In each, the rows were *Ss*, the columns were the order in which the lists were learned, the greek letters were lists, and the roman letters were conditions. To illustrate the assignment of *Ss* to conditions, one of the squares is presented in Table 1.

Experiment II

Except for two changes in procedure, Exp. II was a repetition of Exp. I. The two changes were that the association test was given after the learning of the four critical lists rather than before, as in Exp. I, and the length of the critical lists was extended to 15 items. In all other details, the two experiments were the same.

RESULTS

Practice lists.—Because a detailed analysis of the results of the practice list performances in the two experiments did not shed light on the hypotheses under consideration, the analyses will not be presented. The only significant difference between the R and NR conditions was in the number of correct anticipations on the trial following the criterion of

TABLE 2

MEAN ASSOCIATION TEST VALUES

Cond.	Exp. I				Exp. II			
	Median Latency of First Response (Sec.)		Mean Responses in 30 Sec.		Median Latency of First Response (Sec.)		Mean Responses in 30 Sec.	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
U	2.90	1.05	5.19	2.64	3.42	3.58	4.08	2.46
F	2.96	1.37	5.06	2.56	2.99	1.14	4.00	2.55
	$\sigma_M = .16; t = .40$		$\sigma_M = 1.33; t = 1.04$		$\sigma_M = .34; t = 1.26$		$\sigma_M = .09; t = .89$	

seven correct. Although both conditions exhibited a drop in performance, the drop was significantly greater in the R than in the NR condition.

Association tests.—The association tests for both experiments are summarized in Table 2. For each *S* two median latency scores were obtained: one F score and one U score. The means of these medians are shown in the left column of the table. Similarly, the mean number of responses in 30 sec. to F syllables and to U syllables was determined for each *S*. The means of these values are shown in the right column of Table 2. The values of the *t* tests at the bottom of the table indicate that familiarization had no ascertainable differential effect on either measure of association.

Experiment I critical lists.—Table 3 shows the mean number of trials *S*s used to achieve 7 correct anticipations in one trial, the mean number of trials to 12 correct anticipations, and the mean number of trials from 7 correct to 12 correct. Three analyses of variance were run; one on each of the sets of scores summarized here. The only significant main effect was the effect of familiarization prior to 7 syllables correct ($F = 6.62$; $df = 1$ and 19; $.01 < P < .05$). Neither the effects of the rest interval nor the interaction between rest and familiarization approached statistical significance. Examination of the *SD*s in Table 3 suggests heterogeneity of variance, but analysis of log transformations of these data did not alter conclusions reached from the present analysis.[2] Neither did the rest produce any significant effect on the change in performance from the criterial trial of seven correct to the next trial. These means are shown in Table 4. The *F*s for the effect of rest and the main interaction were both less than 1. The *F* for Familiariza-

TABLE 3

MEANS AND *SD*S OF TRIALS TO INDICATED SYLLABLE CRITERIA IN CRITICAL LISTS: EXP. I

Cond.	Trials to 7 Correct		Trials to 12 Correct		Trials from 7 to 12 Correct	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
NRU	7.50	3.88	20.45	9.27	12.95	8.07
RU	6.90	3.52	18.40	10.02	11.50	7.50
NRF	5.30	2.43	14.90	6.46	9.60	5.48
RF	6.40	1.86	15.60	4.48	9.20	2.25

[2] A search of the literature indicates that procedure for testing for homogeneity of variance in correlated samples has not yet been developed.

TABLE 4

SYLLABLES CORRECT ON 7/12 CRITERION TRIAL AND ON THE NEXT TRIAL IN CRITICAL LISTS: EXP. I

Cond.	Trial 7/12		Trial 7/12+1		Diff.
	Mean	*SD*	Mean	*SD*	
NRU	7.80	.93	7.55	1.47	−.25
RU	7.55	.81	7.00	1.45	−.55
NRF	7.80	.87	7.05	1.56	−.75
RF	7.55	.74	6.70	1.23	−.85

tion was 2.32 ($df = 1$ and 19). Despite this general lack of significance in the main results, two major classes of overt errors were significantly influenced by the familiarization procedure. The mean number of intralist errors per trial increased from .37 (U) to .72 (F), ($P < .01$). The mean number of all other overt errors per trial decreased from .88 (U) to .46 (F), ($P < .01$). Since the major effects were more reliable in Exp. II, a detailed analysis of the errors will be deferred until later.

Experiment II critical lists.—There were two reasons why Exp. II was run with the modifications listed in the method section. First, the effect of familiarization on learning speed was very unreliable. The failure to find a larger effect might have been because the U items became somewhat familiar during the association test. Consequently, the association test was placed after the learning of the critical lists. The purpose of the increase in the length of the critical lists from 12 to 15 syllables was to increase the difficulty of the lists and thereby increase the possibility that some effect of the rest would appear.

The number of trials to criteria of 7 correct, 15 correct, and from 7 correct to 15 correct are shown in Table 5. Familiarization significantly reduced trials to learn in all three cases (trials to 15 correct, $F = 11.38$, $df = 1$ and 19, $P < .01$; trials to 7 correct, $F = 16.11$, $df = 1$ and 19, $P < .01$; trials from 7 correct, $F = 6.51$, $df = 1$ and 19, $.01 < P < .05$). No other experimental manipulation was statistically significant.

Table 6 shows the mean correct responses on the trial on which seven syllables were first correctly anticipated and on the next trial. In the analysis of variance on the change in the number of syllables correct, both the effect of rest and the effect of familiarization resulted in F's of less than 1. However, the interaction of familiarization and rest was significant at the 2% level ($F = 7.54$, $df = 1$ and 19). This significant interaction reflects a beneficial effect of the rest interval that appears in the

TABLE 5

TRIALS TO INDICATED SYLLABLE CRITERIA ON CRITICAL LISTS: EXP. II

Cond.	Trials to 7 Correct		Trials to 15 Correct		Trials from 7 to 15 Correct	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
NRU	8.50	2.84	28.85	10.91	20.35	9.05
RU	8.25	3.05	28.20	11.17	19.95	9.28
NRF	6.40	3.32	21.55	6.39	15.15	5.42
RF	6.05	2.38	22.75	7.62	16.70	6.83

TABLE 6

SYLLABLES CORRECT ON 7/12 CRITERION TRIAL AND ON THE NEXT TRIAL IN CRITICAL LISTS: EXP. II

Cond.	Trial 7/15		Trial 7/15+1		Diff.
	Mean	*SD*	Mean	*SD*	
NRU	7.85	1.06	7.90	1.70	+.05
RU	7.55	.67	6.80	1.83	−.75
NRF	7.90	1.14	7.00	2.07	−.90
RF	7.70	.71	7.30	1.31	−.40

TABLE 7

CLASSIFICATION OF MEAN RESPONSES PER TRIAL ON CRITICAL LISTS INTO FOUR MAJOR GROUPINGS: EXP. II

Cond.	Correct Responses		Omissions		Intralist Errors		Other Overt Errors	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Precriterion 7/15								
NRU	3.94	.79	9.53	1.75	.38	.69	1.17	.94
RU	3.77	.50	9.41	2.07	.43	.81	1.39	1.15
NRF	4.28	.57	9.21	2.06	.93	1.32	.58	.61
RF	4.08	.70	9.44	2.27	.85	1.28	.64	.88
Postcriterion 7/15								
NRU	10.69	1.10	2.87	.98	.32	.33	1.12	.82
RU	10.50	.87	3.07	1.01	.40	.62	1.03	.49
NRF	10.77	.92	3.04	1.00	.67	.70	.52	.35
RF	10.84	.67	2.88	1.03	.69	.58	.60	.36

F conditions but does not appear in the U conditions. It should be noted that the same interaction appears in the mean differences in Exp. 1, but there it is not statistically significant.

The analysis of errors is shown in Table 7. In order to give a complete picture of the effect of familiarization on the nature of the responses, the number of correct responses per trial is also included in the table. The number of correct responses per trial is not significantly influenced by familiarization ($F = 2.91$; $df = 1$ and 19; $P > .05$). Neither is the number of omissions nor, of course, the complement of omissions, the total number of responses per trial ($F = .24$, $df = 1$ and 19). The main effects of the familiarization are to increase intralist errors per trial ($F = 19.36$; $df = 1$ and 19; $P < .01$) and to decrease other overt errors per trial ($F = 28.14$; $df = 1$ and 19; $P < .01$) by roughly corresponding amounts. Table 8 shows a breakdown of these "other" errors per trial into the various subclasses. We have not attempted to analyze the significance of the differences here because of the extremely small values in some cases. However, the major sources of "other" errors are obvious: viz., inventions (made-up syllables not in any practice or critical list) and partial responses. It is equally clear that the reduction in "other" errors as a result of familiarization is directly attributable to a reduction in these two types of errors.

To summarize, the effect of familiarization is two-fold. First reminiscence is produced on the first postrest

TABLE 8

CLASSIFICATION OF MEAN "OTHER" ERRORS PER TRIAL IN CRITICAL LISTS: EXP. II

Cond.	Interlist Intrusions			Inventions	Partial Responses
	From F Lists	From U Lists	From P Lists		
U	.02	.01	.04	.52	.59
F	.06	.01	.02	.13	.35

trial. This beneficial effect is transient, and does not appear in trials to criterion. Second, familiarization reduces the number of erroneous combinations of letters and incomplete responses. In their place, intralist errors tend to occur. While Hovland and Kurtz (1952) found the same reduction in "other" errors, they did not find an increase in intralist errors. Perhaps the most important reason for this discrepancy is that they analyzed mean total errors, whereas we have analyzed mean errors per trial in order to correct for differences in number of trials.

Discussion

The evidence from the present experiments shows that familiarization of syllables does not affect either their association latency or their association frequency. Apparently, the familiarization effect as shown by increased learning speed in the familiarized conditions does not depend on the increased availability of meaningful associates acting as mediators.

To account for the facilitative effect of familiarization in the present experiments, the following facts, as shown in the analysis of Exp. II must be considered. Following familiarization, total response rate remains relatively unchanged, intralist errors per trial increase, and inventions and partial responses per trial decrease.

One possible way to account for these facts is to assume that the response rate is fixed by factors such as instructions, the sophistication of the *S* in the situation, and properties of the *S* such as willingness to guess. If familiarization does not influence response rate, then its main effect, aside from speeding learning, is to bring about a redistribution of errors. In the U condition, *S*s make many partial errors and inventions. During learning, they simultaneously learn what letters go together to form the syllables and where syllables go. Hence there is no great build-up of intralist errors. On the other hand, in the F condition, *S*s, on coming into the serial learning situation, have already learned the letter sequences. Learning is consequently more rapid because only the position of the items must be learned. The fact that the letter sequences have already been learned reduces the tendency to make partial errors and inventions. Given a constant response rate, the inevitable result is a rise in intralist errors. It is inevitable because the familiarized items are strong but their location has not been learned. According to this view, the rise in intralist errors is merely a by-product of the same procedure that reduced trials to learn.

This interpretation is most similar to Hovland and Kurtz' suggestion that the effect of familiarization is to supply *S* with a limited pool of responses from which anticipations are to be made. The increase in intralist errors in the familiarized conditions argues against the notion that the effect of familiarization is to reduce intralist generalization. The small number of interlist errors in both the familiarized and the nonfamiliarized conditions shows that discrimination between the two types of items, as measured by interlist intrusions, is at least not a necessary condition for the facilitative effect of familiarization.

The presence of reminiscence in the familiarized conditions of Exp. II points to the condition that may be necessary for producing this highly elusive phenomenon, namely, a high degree of interference between items. The problems that this sort of finding raises for a general interpretation of distribution effects have been discussed by Riley (1957) in another paper.

Summary

Two experiments were conducted that evaluated the effects of nonsense syllable familiarization on: (*a*) the association value of the syllables, (*b*) the speed of learning serial lists, and (*c*) performance following a 2-min. color-naming interval after seven correct anticipations. Trained *S*s were run under all conditions.

Although familiarization had no effect on association value, it did facilitate learning. Familiarization also produced both reminiscence and a marked increase in intraserial errors.

References

GIBSON, E. J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.*, 1940, **47**, 196–229.

GLAZE, J. A. The associative value of nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 255–269.

HOVLAND, C. I. Experimental studies in rote learning theory. I. Reminiscence following learning by massed and distributed practice. *J. exp. Psychol.*, 1938, **22**, 201–224.

HOVLAND, C. I., & KURTZ, K. H. Experimental studies in rote learning theory: X. Pre-syllable familiarization and the length-difficulty relationship. *J. exp. Psychol.*, 1952, **44**, 31–39.

HULL, C. L. The meaningfulness of 320 selected nonsense syllables. *Amer. J. Psychol.*, 1933, **45**, 730–734.

MANDLER, G., & HUTTENLOCHER, J. The relation between associative frequency, associative ability and paired-associate learning. *Amer. J. Psychol.*, 1956, **69**, 424–428.

MELTON, A. W., & STONE, G. R. The retention of serial lists of adjectives over short time-intervals with varying rates of presentation. *J. exp. Psychol.*, 1942, **30**, 295–310.

NOBLE, C. E. An analysis of meaning. *Psychol. Rev.*, 1952, **59**, 421–430. (a)

NOBLE, C. E. The role of stimulus meaning (*m*) in serial verbal learning. *J. exp. Psychol.*, 1952, **43**, 437–446. (b)

NOBLE, C. E. The effect of familiarization on serial verbal learning. *J. exp. Psychol.* 1955, **49**, 333–337.

RILEY, D. A. The influence of amount of prerest learning on reminiscence effects in paired-associate learning. *J. exp. Psychol.*, 1957, **54**, 8–14.

UNDERWOOD, B. F., & RICHARDSON, J. Studies of distributed practice: XIII. Interlist interference and the retention of serial nonsense lists. *J. exp. Psychol.*, 1955, **50**, 39–46.

WARD, L. B. Reminiscence and rote learning. *Psychol. Monogr.*, 1937, **49**, No. 4 (Whole No. 220).

(Received June 16, 1958)

Journal of Experimental Psychology
Vol. 57, No. 6, 1959

THE EFFECTS OF MOTIVATION AND PROBABILITY OF REWARD ON TWO-CHOICE LEARNING [1]

PAUL J. WOODS

Hollins College

In the present experiment three variables are investigated in relation to a two-choice task with contingent reward. The variables are motivation, in the form of an aversive auditory stimulus, correction vs. noncorrection procedure, and probability of reward. The results will be interpreted in terms of Estes' statistical learning theory (Estes, 1950, 1954; Estes & Straughan, 1954).

METHOD

Design.—The schedules of probability of reward were 98:58, 87:47, and 53:13,[2] which respectively predict the following asymptotes in Estes' theory: .95, .80, and .65. The notation used for the reward schedules indicates the percentage of times reward followed each alternative, and the notation used for the asymptotes refers to the proportion of times the theory predicts a group of *S*s will ultimately choose the alternative with the higher probability of reward (π_1). The event probabilities were randomized separately for each *S* within blocks of 20 trials, so successive blocks contained the approximate proportions of the particular schedule. For the entire series of trials, however, the probability schedule was attained.

Each schedule was used under a correction and a noncorrection procedure and three conditions of motivation. The motivation was varied by means of the intensity of a noxious auditory stimulus produced by a noise generator. The low intensity noise, as measured by a sound pressure level meter, was 70 ± 3 db, and the high intensity noise was 95 ± 3 db. The medium intensity was chosen at 80 ± 3 db on the basis of data from five *S*s who estimated this level to be halfway between the low and high intensity noises. Conse-

TABLE 1

ANALYSIS OF VARIANCE OF THE PROPORTIONS OF CHOICE TO THE MORE REWARDED SIDE IN BLOCKS OF 20 TRIALS

Source	*df*	*MS*	*F*
Main-plot analysis			
Procedure† (P)	1	0.034	0.2
Schedule (S)	2	5.725	36.6*
P × S	2	0.033	0.2
Motivation (M)	2	0.152	1.0
P × M	2	0.180	1.2
S × M	4	0.135	0.9
P × S × M	4	0.050	0.3
Error	90	0.156	
Total	107		
Subplot analysis			
Blocks (B)	14	0.847	79.9*
B × P	14	0.010	0.9
B × S	28	0.050	4.8*
B × M	28	0.014	1.4
B × P × S	28	0.010	1.0
B × P × M	28	0.016	1.5
B × S × M	56	0.010	1.0
B × P × S × M	56	0.014	1.3
Error	1260	0.011	
Over-all Total	1619		

* $P < .001$.
† Correction vs. noncorrection.

[1] The present paper is adapted from a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the University of Connecticut. The writer wishes to express appreciation to his advisor, Howard Brand, for advice throughout the course of the research. Indebtedness to James M. Sakoda and David Zeaman for their many helpful suggestions is also acknowledged.

This research was supported, in part, by the National Institute of Mental Health under Grant M-814.

[2] The reader should note that the difference between the probabilities of reward is constant, but the ratios are not so controlled. Brand, Woods, and Sakoda (1956) have found that with a constant ratio, increasing the difference produced increasing asymptotic levels, but with a constant difference, increasing the ratios produced no consistent influence on the asymptotic levels.

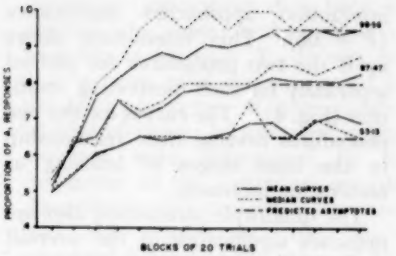

FIG. 1. The mean and median proportions of choices to the high π alternative for each schedule (pooled over conditions of motivation and procedure).

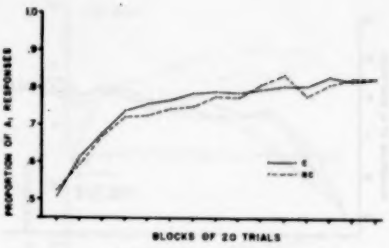

FIG. 2. The mean proportions of choices to the high π alternative for the correction and noncorrection procedure (pooled over schedules and conditions of motivation).

quently the physical stimuli were at approximately equal psychological intervals. Intensities greater than 95 db were not used for fear of damaging *Ss*' ears.

All possible combinations of the three independent variables gives 18 experimental conditions. Six *Ss* were assigned randomly to each condition, and they were given 300 consecutive acquisition trials.

Apparatus.—A two-choice contingent reinforcement apparatus was used which has been fully described elsewhere (Brand, Sakoda, & Woods, 1957). The only modification was the addition of a noise generator whose output was fed into an ordinary amplifier and delivered to *S* through a headset.

Subjects.—The *Ss* were 108 undergraduate, unpaid volunteers at the University of Connecticut. There were 74 women and 34 men. The *Ss* were asked if they had any auditory impairment; none reported any.

Procedure.—Details are available elsewhere (Woods, 1957). The task presented *S* was that of trying to outguess *E* on each trial by always pressing the switch he thought would give the red light rather than the white light. Then *E* set the electrical connections for one of the probability schedules, and *S* was given the green ready signal and the aversive stimulus. For the noncorrection procedure, if *S* chose incorrectly, a white light flashed and the noise continued. For the correction procedure, if *S* chose incorrectly, he was required to make the opposite response. The correction response was followed by a red light and termination of the noise, as were correct responses in either procedure. The interval between trials was about 2 sec.[3]

[3] A criticism of this procedure might be that *S* is responding only to the instructional task of trying to get red lights, and hence it is not certain that the noise alone is capable of influencing behavior. To answer this criticism a small study was conducted with a single schedule of reinforcement (87:47), nine *Ss* being run under each noise intensity without any reinforcement lights. All *Ss* learned to prefer the noise terminating alternative and the mean curve did not differ significantly from the 87:47 curve of the main experiment.

RESULTS

Over-all analysis.—A split-plot analysis of variance was performed on the proportions of responses to the high π alternative in blocks of 20 trials and is summarized in Table 1. First the main effects will be discussed and then the higher-order interactions will be given.

Schedules of probability of reward.—Figure 1 shows a plot of the mean proportions of responses to the high π alternative in blocks of 20 trials for each schedule. The differential effect of the schedule of reward probabilities is so strong that only in the early learning trials do two of the curves intersect.

Procedure (correction and noncorrection).—The mean proportions of choices to the most rewarded alternative for the correction and noncorrection procedure are shown in Fig. 2. There is no significant difference

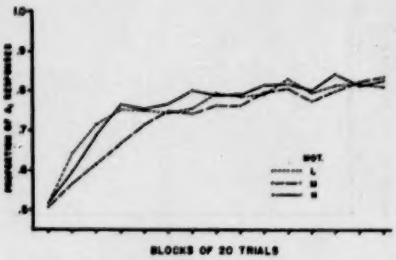

FIG. 3. The mean proportions of choices to the high π alternative for conditions of motivation (pooled over schedules and procedure).

although for 10 out of the 15 points, the correction curve is above the noncorrection curve.

Motivation.—The mean proportions of responses for each of three levels of motivation are plotted in Fig. 3. The curve for the intermediate level of motivation is initially below those for the low and high conditions. Such a relationship suggests a nonlinear trend between motivation and the rate of learning. No significant difference was found for these conditions, however, in the analysis of variance which tested for mean level differences or in the analysis by orthogonal polynomials (Grant, 1956) which tested for differences in trend.

To investigate further the possibility of a nonlinear trend between motivation and the rate of learning, one can use the rate constants from Estes' prediction equation (Brand et al., 1957) as a measure of the rate of learning. The values based on empirical asymptotes for each of the 18 conditions are shown in Table 2. It may be noted that the relation between the rate constant and motivation is U shaped in five of the six instances.

Higher-order interactions.—In the over-all analysis of variance the Blocks × Procedure × Motivation interaction approaches significance ($P = .06$). This interaction shows up if the two procedures are plotted separately for each motivating condition (Fig. 4). The curves for the two procedures reverse their relationship in the later stages of learning as motivation increases.

The quadruple interaction also approaches significance in the over-all analysis of variance ($P = .07$).

Asymptotic levels.—The theory predicts asymptotes by means of the following formula which has been independently derived by Estes (1954) and Bush and Mosteller (1955, p. 288):

$$\frac{1 - \pi_2}{2 - \pi_1 - \pi_2}$$

where π_1 and π_2 are the probabilities of reward on the more rewarded and the less rewarded alternative, respectively. As shown in Fig. 1 the lower curve (53:13) rises considerably above the predicted asymptote; the middle curve (87:47) is just slightly above it; and the upper curve (98:58) is a bit below it. It was felt that by plotting frequency distributions of individual asymptotic estimates under each schedule a better picture of these relationships might be obtained. A technique that has been used for estimating

TABLE 2

RATE CONSTANTS ($\theta(2 - \pi_1 - \pi_2)$)

Schedule	Procedure	Motivation		
		Low	Middle	High
98:58	Cor.	.018	.015	.026
	Noncor.	.024	.011	.021
87:47	Cor.	.036	.013	.019
	Noncor.	.014	.015	.017
53:13	Cor.	.015	.012	.013
	Noncor.	.016	.010	.016

individual asymptotes has been to average the same final number of trial blocks for each *S*. It was felt that a more valid estimate could be obtained, however, if the number of trial blocks to be averaged in this estimate were determined by an inspection of the individual records. Accordingly this technique was used and the number of trial blocks averaged fell within the range of 5 to 11 for over 80% of the *S*s. For those records which did not appear to be leveling off, the "asymptote" was computed by averaging the final two or three blocks.

The distributions were markedly skewed in all three cases. The *t* tests comparing the means of the individual values with the predicted values gave a significant discrepancy for the 53:13 schedule ($P < .01$). An analysis of variance of these individual asymptotic estimates indicated that the schedules produced the only significant differences ($P < .0005$). A point of importance here is that motivation in this task does not markedly raise the asymptotic level.

To obtain a more sensitive analysis of the relationship in the later parts of the records, split-plot analyses of variance were carried out on the final eight trial blocks under each motivating condition. The only finding of interest is that under low motivation, trials are not significantly different; they are highly significant under middle motivation ($P < .005$); and they are not significant under high motivation. There is, therefore, a U-shaped relationship between motivation and the stage of learning at which a stable final response level is achieved. This finding was reflected in the rate constants shown in Table 2.

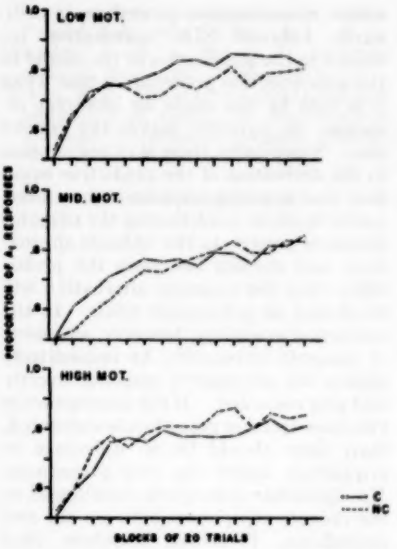

FIG. 4. The mean proportions of choices to the high π alternative for the correction and noncorrection procedures under each condition of motivation (pooled over schedules).

DISCUSSION

Schedules.—Increasing the over-all probability of reward, when the difference between π values for the two alternatives is constant, produces significantly increasing proportions of choices to the more rewarded alternative: a result that has already been found in previous studies. Analysis of the data obtained here indicates that with schedules leading to high asymptotic levels we obtain very skewed distributions of empirical points in the later stages of learning. Consequently, mean curves can profitably be supplemented by additional data, such as the median curves plotted in Fig. 1, to give more information concerning the course of two-choice learning in groups of *S*s.

Procedure.—Two experimental procedures were used in this experiment to test an assumption in Estes' model for the contingent reinforcement situation,

where noncorrection procedure is ordinarily followed. The assumption involved in the application of the model to the noncorrection procedure is that when *S* is told he has made an incorrect response, he covertly makes the correct one. Specifically, there is an assumption in the derivation of the prediction equation that a wrong response to one alternative leads to conditioning the stimulus elements present to the *opposite* alternative, and thereby increases the probability that the opposite alternative will be chosen on subsequent trials. In the correction procedure, however, whenever *S* responds incorrectly, he immediately makes the alternative response overtly and gets rewarded. If the assumption in the noncorrection procedure is warranted, then there should be no difference in acquisition under the two procedures. No significant differences were found in the present experiment between the two procedures. It seems, therefore, that this assumption is tenable, at least when other variables which may possibly be producing interaction effects are pooled.

The Trials × Procedure × Motivation interaction approaches significance, however, and the results as plotted in Fig. 4 do merit consideration. In the final half of the records, the acquisition curves for the two procedures completely reverse their relative positions as motivation increases. It is possible that real effects are being obscured here and elsewhere by the large amount of variation among *S*s in the same experimental conditions.

Motivation.—Motivation in the form of a noxious auditory stimulus had no unequivocal effect on the rate of learning. Several aspects of the data suggest a curvilinear relationship. It is interesting to note, however, that this U-shaped relationship is just the opposite of the inverted U-shaped relationship between motivation and efficiency which has been mentioned in many textbooks. (See, for example, Wickens & Meyer, 1955, p. 137.) Of course, the noxious stimulus in the present experiment may have influenced certain aspects of *S*'s behavior which were not measured, such as the speed of responding. And further, speculating on an *ad hoc* basis, it may have stimulated an aspect of behavior which usually leads to greater efficiency, viz., ideas or "hypotheses" about the problem at hand. This very stimulation in this sort of task does not necessarily lead to a greater proportion of choices to the high π alternative. It leads *S* to try various hypotheses, such as notions about the patterning of reinforcements, etc., and thereby increases the number of choices to the low π alternative. This may then be the reason for the U-shaped relationship suggested by the present experiment.

Summary

The purpose of the experiment was to investigate the relationship between behavior in a two-choice learning task and three experimental conditions: motivation, in the form of an aversive auditory stimulus, correction and noncorrection procedure, and schedule of contingent reward. In all there were three levels of motivation, two experimental procedures, and three schedules of probability of reward. The acquisition data of the six *S*s run under each of the conditions were compared and the results were further analyzed in the context of Estes' statistical learning theory. The main effect was due to the probability of reward. The other variables enter into the process to a lesser degree. The data suggest the relationship between motivation and rate of learning may be U-shaped in nature.

References

Brand, H., Sakoda, J. M., & Woods, P. J. Contingent partial reinforcement and the anticipation of correct alternatives. *J. exp. Psychol.*, 1957, **53**, 417–424.

Brand, H., Woods, P. J., & Sakoda, J. M. The anticipation of reward as a function of partial reinforcement. *J. exp. Psychol.*, 1956, **52**, 18–22.

Bush, R. R., & Mosteller, F. *Stochastic models for learning.* New York: Wiley, 1955.

Estes, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, **57**, 94–107.

Estes, W. K. Individual behavior in un-

certain situations: An interpretation in terms of statistical association theory. In R. M. Thrall, C. H. Coombs, & R. L. Davis (Eds.), *Decision processes*. New York: Wiley, 1954.

ESTES, W. K., & STRAUGHAN, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.*, 1954, **47**, 225–234.

GRANT, D. A. Analysis-of-variance tests in the analysis and comparison of curves. *Psychol. Bull.*, 1956, **53**, 141–154.

WICKENS, D. D., & MEYER, D. R. *Psychology*. New York: Dryden, 1955.

WOODS, P. J. The effects of motivation and probability of reward on discrimination learning. Unpublished doctoral dissertation, Univer. of Connecticut, 1957.

(Received June 23, 1958)

Journal of Experimental Psychology
Vol. 57, No. 6, 1959

AGE DIFFERENCES IN THE ACQUISITION AND EXTINCTION OF THE CONDITIONED EYELID RESPONSE [1]

HARRY W. BRAUN AND RICHARD GEISELHART

University of Pittsburgh

This study was stimulated in part by a report of Gakkel and Zinina (1953) that *S*s over 60 years of age required a longer period for conditioning of the eyeblink and also for its extinction. It was also performed to provide systematic information on the conditioning and extinction performance of children, young adults, and old adults under the same controlled conditions. Apparently this problem has been hitherto unexplored.

METHOD

Subjects.—These were 15 boys in the age range 8–10 yr., with a mean age of 9.36 yr.; 15 male young adults (college students) aged 18–25 yr. whose mean age was 20.63 yr.; and 13 male adults whose mean age was 70.5 yr., range 62–84 yr. The old adults were noninstitutionalized and were either employed full time or part time, or retired. All *S*s were volunteers and some of the elderly group were compensated for their services.

Apparatus and method of recording.—The equipment for recording the eyelid closure and for presenting the CS and UCS was similar to that used by Spence (1953), and will not be described here. The *S* sat in a dental chair in a darkened room adjoining that in which the stimulus controls and recording apparatus were located. He was instructed to blink on receiving a ready signal, and then to fixate on a 6-cm. circular milk glass disc placed at a distance at 125 cm.

The CS was an increase in the brightness of this disc from 0.05 to 1.5 apparent ft.-candles. The duration of the CS on each trial was 1 sec. The UCS was a puff of air with an intensity of 2.0 lb./sq. in. delivered to the right eye, with a duration of 500 msec. The CS-UCS interval was 500 msec. All *S*s were given 80 conditioning trials, on each of which the light and air puff were paired. Intertrial intervals of 10, 15, and 20 sec. were used according to a prearranged schedule. Following the last conditioning trial, 20 extinction trials were given, in which both light and air puff were paired, but in which the CS-UCS interval was lengthened to 1500 msec. Only those CR's whose latency was 300 msec. or more were included in the data. No *S*s were excluded as voluntary responders, i.e., *S*s who gave at least half of their CR's with a latency of less than 300 msec.

RESULTS

Curves of acquisition of the CR for the three age groups are shown in Fig. 1 in terms of the percentage of CR's in each block of 10 trials. Through the 60th conditioning trial, the curve for the children is consistently above that of the young adults, and it is above that of the elderly group throughout the entire 80 trials. The conditioning performance of the young and old adults is similar through the first 20 trials after which

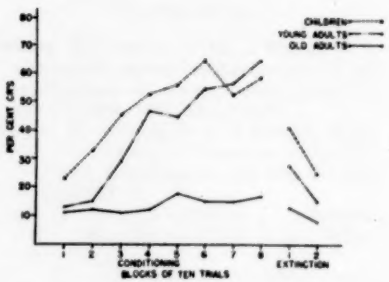

FIG. 1. Conditioning and extinction as a function of age.

[1] This investigation was supported in part by a research grant, M-1365, from the National Institute of Mental Health, U. S. Public Health Service.

the young group improves markedly and the elderly group only slightly. The conditioning curve of the children is negatively accelerated, that of the young adult group is positively accelerated, while the curve of the old adults is flat. At the start of conditioning, the performance level of the children is relatively high and/or the rate of conditioning is high. On the other hand, the performance level of the other two age groups at the beginning of conditioning is relatively low and/or the rate of conditioning is slow.

Statistical evaluation of these results was made by the Mann-Whitney test. The difference between the number of CR's made by the children and the old adults was significant beyond the .002 level for a two-tailed test ($U = 23$) and the difference between the number of CR's made by the young adults and the old adults was also significant beyond the same level of confidence ($U = 25$). The children and young adults did not differ significantly in the number of CR's ($U = 89$).

Curves of latency of the CR's in blocks of 20 trials for the three age groups appear in Fig. 2. These data, which are measures of latency of CR to the onset of the CS on the 80 acquisition trials, show that during the first 60 trials, the CR's of the children are slower than those of the other two groups. During the last 20 conditioning trials, however, the three age groups do not appreciably differ on this measure. The sharp decrease in latency shown by the elderly group in the third block of twenty trials was caused by two new responders.

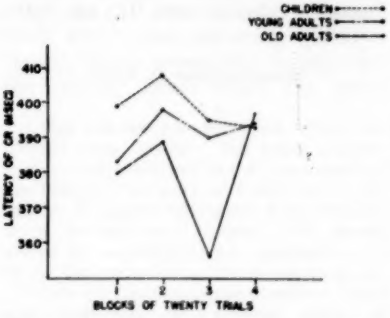

FIG. 2. Latency of CR as a function of age.

The use of the Mann-Whitney test showed that the children made significantly more CR's during extinction than the elderly adults ($U = 27$; $P < .002$) as did the young adults ($U = 49$, $P = .05$). The two younger groups did not differ significantly in this respect. However, after analysis of covariance, age differences in extinction were not significant.

DISCUSSION

The main and striking finding of this study is the relative inability of the elderly *Ss* to acquire the conditioned eyeblink response. Of the 13 *Ss* in this group, four *Ss* gave no CR's during the 80 conditioning trials, seven gave from one to eight CR's, while 37 and 75 CR's were given by the other two *Ss*. There were no instances of failure to condition in the two younger groups. This result confirms the report of Gakkel and Zinina (1953) although in their study data on the conditioning performance of younger *Ss* were not included. To account for the failure of older *Ss* to acquire CR's, Gakkel and Zinina quote Pavlov as coming to the conclusion that "liveliness of the nerve processes suffer(s) from the development of senile changes." They also invoke the construct of "inertia" of the stimulating process at senile age.

However, several considerations indicate that this hypothesis may not be entirely satisfactory. The latency of the CR's of the elderly group are not longer than those of the other groups. It is admitted that these are tenuous data since they involve only a very small

number of responses. However, Weiss (1956) has shown that the latency of the retina as reflected in the retinogram does not change with age. His *Ss* were 15 males aged 66 to 76 yr. and 10 between the ages of 18 to 37 yr. Similarly, Kumnick has reported that "the efficiency of the pupillary mechanism in relation to similar original diameter in the response to light stimuli did not decrease with increasing age" (Kumnick, 1956, p. 160). Thus, activation time of at least these aspects of the visual organ seems to be unchanged with age. It would be premature to conclude that the latencies of all simple involuntary responses or reflexes do not change with age. Magladery, Teasdall, and Norris (1958) have shown that the latencies of plantar flexor responses and superficial abdominal reflexes are markedly slower in elderly *Ss* and children than in young adults.

The authors suggest the consideration of an adaptation hypothesis to account for the relative unconditionability of elderly *Ss*. Both MacDonald (1946) and Taylor (1956) have reported that *Ss* who were adapted to the UCS (air puff) show a significantly lower level of conditioned eyelid responses than *Ss* who were not given preadaptation trials. It is proposed that in the course of many years of living, the eyelid response as well as probably other responses have been "adapted out" and thus less susceptible to subsequent conditioning. Research is under way to study several consequences of this hypothesis: that both frequency and amplitude of blink are reduced as a function of age. The procedures employed in the present study did not provide data on these variables.

Summary

This study investigated eyelid conditioning and extinction of children, young adults, and old adults under the same controlled conditions. All *Ss* were given 80 conditioning trials and 20 extinction trials. The children and young adults gave significantly more conditioned responses than the old adults. Elderly *Ss* were relatively unable to acquire the conditioned eyeblink response. Latency of response and extinction were not related to age.

References

Gakkel, L. B., & Zinina, N. V. Changes of higher nerve function in people over 60 years of age. *Fiziolog. Zhurnal.*, 1953, **39**, 533–539.

Kumnick, L. S. Aging and the efficiency of the pupillary mechanism. *J. Geront.*, 1956, **11**, 160–164.

MacDonald, A. The effect of adaptation to the unconditioned stimulus upon the formation of conditioned avoidance responses. *J. exp. Psychol.*, 1946, **36**, 1–12.

Magladery, J. W., Teasdall, R. D., & Norris, A. N. Effect of aging on plantar flexor and superficial abdominal reflexes in man—a clinical and electromyographic study. *J. Geront.*, 1958, **13**, 282–288.

Spence, K. W. Learning and performance in eyelid conditioning as a function of the intensity of the UCS. *J. exp. Psychol.*, 1953, **45**, 57–63.

Taylor, J. A. Level of conditioning and intensity of the adaptation stimulus. *J. exp. Psychol.*, 1956, **51**, 127–130.

Weiss, A. D. The relation of A-wave latency of the electroretinogram to human aging. *J. Geront.*, 1956, **11**, 448. (Abstract).

(Received June 23, 1958)

Journal of Experimental Psychology
Vol. 57, No. 6, 1959

SOME PHYSIOLOGICAL CORRELATES OF HUNGER AND PAIN [1]

BERNARD T. ENGEL [2]

University of California, Los Angeles

This study is an attempt to identify several specific response patterns mediated by the autonomic nervous system (ANS) in the same group of *S*s. The test conditions were: (*a*) graded periods of food deprivation, and (*b*) graded durations of foot immersion in iced water. The iced water condition was chosen because of its well-known pain-producing effect.

It has been shown that regions exist in the hypothalamus which are related to eating behavior. First to be described was the so-called satiety center (Brobeck, Tepperman, & Long, 1942) in the medial hypothalamus; and, subsequently a region less clearly demarcated in the lateral hypothalamus has been characterized as the hunger center (Anand, Dua, & Shoenberg, 1955). Since the hypothalamus has also been identified as an integrating center for responses mediated through the ANS, it would seem reasonable to expect different ANS patterns related to varying periods of food deprivation.

Several studies have clearly shown that painful stimuli have autonomic effects (e.g., Cannon, 1929; Wolff & Wolf, 1951); and one study (Wolf & Hardy, 1941) attempted to separate the cooling effect from the pain-producing effect of the cold-pressor test.

[1] This research was supported, in part, by USPHS Grant M-788. This paper is taken from a thesis submitted to the Department of Psychology, University of California at Los Angeles, in partial fulfillment of the requirements for the Ph.D. degree. The author wishes to acknowledge his appreciation to M. A. Wenger for his assistance and guidance.

[2] Now at Cardiovascular Research Institute, University of California School of Medicine, San Francisco 22, California.

METHOD

Independent variables.—Four levels of hunger and of pain were investigated. Degrees of hunger were produced by depriving each *S* of food for 2, 10, 14, and 24 hr., the first being regarded as a control period. Ten hours of deprivation was chosen because a shorter period would have interfered with the night's sleep of some *S*s (all experiments were conducted during the morning), and a longer period would have been too similar to the 14-hr. deprivation. The latter was chosen because of its convenient relationship to the normal dinner hour.

Pain was produced by means of the cold-pressor test. This test requires that *S* immerse an extremity (in this case the right foot, approximately to the level of the internal malleolus) into a container of iced water (in this study the temperature was maintained between 3° and 4° C.). Varying levels of pain supposedly were produced by varying the immersion time; zero (the control period), 2, 3, and 4 min. were used.

In order to relate the intensity of the stimuli to subjective impressions, each *S* was required to rate each stimulus in two ways. First, he rated it after each session by designating the stimulus intensity as zero, mild, moderate, or severe. Next, after the second, third and fourth sessions of each stimulus he was asked to compare the stimuli among themselves and to order them in increasing magnitude.

Dependent variables.—The apparatus used in this study has been fully described by Wenger, Engel, & Clemens (1957), and Wenger, Henderson, & Dinning (1957). Table 1 summarizes the functions recorded, units of measurement, sampling frequency, and method of assay.

Experimental design.—In order to counterbalance the stimulation periods, latin square designs were used. There were four squares for the hunger conditions and four separate squares for pain. Each of the four squares within each set was randomly selected; and *S* was randomly assigned to the rows of the

TABLE 1

AUTONOMIC VARIABLES, UNITS AND METHODS OF MEASUREMENT

Function	Unit of Measurement	Recording Frequency	Method of Assay	
			Hunger	Pain
Systolic and diastolic blood pressures and pulse pressure	mm. Hg.	Once per min.	Average of last 5 readings	Average of last 2 readings
Electrical palmar skin resistance (conductance)	ohms (micromhos)	Continuously	Mean level attained based on samples taken at the same times as the blood pressure samples	
Finger, face, axillary and room temperatures	°C	Finger twice and face and axillary temperatures once every 10 sec. and room temperature at beginning of recording session	Mean level attained based on samples taken at the same times as the blood pressure samples	
Heart rate	cycles/min.	Continuously	Mean rate during last 5 min. of recording session	Mean rate during last 2 min. of recording session
Breathing rate	cycles/min.	Continuously	Same as heart rate	Same as heart rate
Finger pulse volume	microliters/stroke/min.	Continuously	Mean of 5 modal amplitudes during last 5 min. of recording session	Mean of 2 modal amplitudes during last 2 min. of recording session
Stomach motility	Presence or absence; sec/cycle of Type I waves	Continuously	Mean period of waves (when present) throughout entire recording session	Mean period of waves (when present) before and during stimulation
Rate of salivary flow	microliters/min.	Continuously	Mean rate during last 5 min. of recording	Mean rate during last 2 min. of recording

squares. In order to reduce the probability of a significant "order" effect, sophisticated *Ss* were used. Since Wenger (1948) has shown certain autonomic variables and time of day are correlated, this effect was confounded with the rows (*Ss*) and the experimental times restricted to three morning hours (9:00, 10:00, or 11:00 A.M.). Each *S* was assigned a given sequence of pain and of hunger and the two sets of situations were interspersed at random. If B, A, C and D represent the sequence of cold-pressor tests, and 4, 3, 1 and 2 the sequence of deprivation sessions for a given *S*, his run might be: B-A-4-C-3-D-1-2. In order to control for any cumulative effects, at least 72 hr. intervened between any two recording sessions.

Each experimental session covered approximately 45 min. Twenty min. were required to connect the leads to *S* and make the necessary equipment adjustments. There followed 10 min. of recording during rest, and 5 min. of critical recording. (In the case of pain, only the last 2 min. were analyzed.) The 10-min. period was designed to allow *S* to adapt to the experimental environment and the data it afforded were not analyzed except in the case of stomach motility (Table 1).

Subjects.—The *Ss* were 16 male psychology graduate students ranging in age from 23 to 40; the mean age was 28.3 yr.

RESULTS

Hunger.—The rating scales were used to determine the relationship between periods of deprivation and subjective impressions, and to determine whether there existed a hierarchical order of subjective sensation corresponding to the deprivation periods. The concept of hunger was never specified for *S*. When he asked for a definition, he was told to use whatever criteria he considered relevant in light of his own experiences.

The coefficient of concordance was chosen to give an index of the degree of consistency among *Ss*' rankings of

the deprivation periods. In order to reduce the two sets of ranks to a single matrix, they were added, re-ranked, and the ties broken by coin tosses. The mean ranks for the 2-, 10-, 14-, and 24-hr. deprivation conditions were: 1.1, 2.2, 3.0, and 3.7, respectively. The coefficient of concordance, .631, was significant beyond the .01 level; and since the difference between any consecutive pair of deprivation period means was approximately the same, the conditions are probably graded in intensity. Taken together, these results are the bases for the conclusion that the increases in deprivation period used in this study produced increments in hunger, when the latter is defined as a sensation.

The effects of food deprivation on each of the physiological functions tested are reproduced in Table 2. Table 3 shows the results of the analyses of variance for those physiological variables showing significant changes.

TABLE 2

MEAN LEVELS OF MEASURED ANS VARIABLES UNDER FOUR FOOD DEPRIVATION CONDITIONS

Variable[a]	Deprivation Period			
	2	10	14	24
Systolic blood pressure	121.0	118.0	118.8	121.4
Diastolic blood pressure	79.5	80.1	80.6	80.1
Pulse pressure	41.5	37.9	38.9	41.3
Heart rate	68.1	66.5	67.6	68.0
Rate of salivary flow	348	280	354	441
Palmar conductance (log × 10²)	672	623	666	708
Respiration rate	17.2	16.6	16.5	15.8
Face temperature	34.8	34.7	34.9	35.2
Finger pulse volume	2.3	2.6	2.9	3.4
Axillary temperature	36.2	36.3	36.5	36.4
Finger temperature	32.4	31.2	31.8	32.3

[a] Cf. Table 1 for units of measurement.

An analysis of room temperature fluctuations was included because previous studies have shown that they can influence some autonomic variables (Wenger, 1948). The mean room temperatures during the deprivation sessions ranged from 24.1° C. to 24.4° C.; the F was insignificant.

TABLE 3

ANALYSES OF VARIANCE OF DATA COLLECTED UNDER FOUR CONDITIONS OF FOOD DEPRIVATION[a]

Source	*df*	Blood Pressure				Salivary Flow[b]		Breathing Rate		Axillary Temp.	
		Systolic		Pulse							
		MS	*F*	*MS*	*F*	*MS*	*F*	*MS*	*F*	*MS*	*F*
Food Deprivation	3	43.61	2.25	51.83	3.11*	67.6	2.86*	5.17	3.42*	.273	2.18
Linear	1	3.14	—	.22	—	109.9	4.66*	14.66	9.71**	.612	4.90*
Quadratic	1	124.04	6.39*	146.41	8.78**	73.6	3.12	.01	—	.051	—
Cubic	1	3.63	—	8.84	—	19.4	—	.85	—	.153	1.22
Order	3	33.92	1.75	31.47	1.89	42.8	1.81	6.10	4.04*	.037	—
Ss	15	192.53	9.91***	135.21	8.11***	220.3	9.33***	96.98	64.22***	.231	1.85
Error	42	19.41		16.68		23.6		1.51		.125	
Total	63										

[a] The analyses of variance for all functions appeared in Vol. 4, No. 1, 1958 of the *Psychophysiology Newsltr.* Dittoed copies of the raw data may be obtained by writing to: Albert F. Ax, Scribe; *Psychophysiology Newsltr.*; Lafayette Clinic; 951 E. Lafayette; Detroit 7, Michigan.

[b] Gap between the 24-hr. fast and the 10-hr. fast significant beyond the .01 level; gap between the 24- and 2-hr. fasts significant beyond the .05 level; gap between the 10- and 14-hr. fast approaches significance at the .05 level (gap = .83; R_p = .843).

* P = .05.
** P = .01.
*** P = .001.

The F's for pulse pressure, respiration rate, and salivary output are statistically significant (.05 level or greater); however, only the means for the last showed significant gaps when Duncan's test (1955) was applied. Twenty-four hours of deprivation produced greater salivary flow than either 10 or 2 hr., and the 14-hr. deprivation condition produced more salivary flow than did the 10-hr. deprivation. Some of the original data were markedly skewed and required correction before analyses of variance tests could be applied. Salivary outputs and finger pulse volumes were transformed to square roots, skin resistances to log conductances, and finger temperatures to log (37°-finger temperature). In the last case, the data had to be translated because they were negatively skewed. All transformations were determined on the basis of visual inspection of the frequency plots.

In all cases in which the sums of squares for deprivation were at least four times as great as the mean square for error (for 1 and 40 df, $F_{.05} = 4.08$), the deprivation effects were tested to determine whether they had statistically significant linear, quadratic, and cubic components (Table 3). In two instances (systolic blood pressure and axillary temperature) the trend components were significant when the deprivation effects were not. While data such as these must be regarded with some skepticism, they cannot be ignored. Both systolic blood pressure and pulse pressure have significant quadratic trends, whereas the other three have linear components. The salivation data may have a quadratic component, however, since the gap between the 10- and 14-hr. deprivation periods, .83, falls just short of significance at the .05 level (.843), and the gap between the 2- and 10-hr. periods, .74, approaches significance ($R_{.05} = .801$). The F for quadraticity, 3.12, is well short of that required for significance ($F_{.05} = 4.08$).

TABLE 4

MEAN LEVELS OF MEASURED ANS VARIABLES DURING FOUR EXPOSURES TO ICED WATER

Variable[a]	Exposure Time (Min.)			
	0	2	3	4
Systolic blood pressure	125.8	134.1	138.0	136.8
Diastolic blood pressure	81.1	92.8	93.6	90.9
Pulse pressure	44.6	41.2	44.4	45.8
Heart rate	69.6	77.1	71.3	74.5
Rate of salivary flow	375	262	339	279
Log conductance ($\times 10^6$)	722	766	772	841
Respiration rate	16.8	17.5	17.8	18.0
Face temperature	35.2	35.1	34.8	35.0
Finger pulse volume	2.6	1.5	1.7	2.0
Axillary temperature	36.2	36.3	36.4	36.3
Finger temperature	32.4	32.0	31.3	31.4

[a] Cf. Table 1 for units of measurement.

The data for stomach motility were not analyzed statistically, since their form precludes the use of most standard statistical procedures. In the 2-, 10-, 14-, and 24-hr. deprivation conditions 9, 7, 10, and 11 *S*s showed Type 1 waves,[3] respectively, and the mean periods of these waves were 19.9, 19.4, 19.5, and 19.0 sec., respectively.

It can be seen that the incidence of activity increased with increasing food deprivation except for the 10-hr. fast; and that the mean period of contraction decreases as deprivation time

[3] The most commonly accepted terminology for designating stomach contractions according to type is that given by Code, Hightower, and Morlock (1952). Type I waves are those with periods ranging from about 5 to 20 sec.; Type II waves range in period from 12 to 75 sec.; and Type III waves cover about 3 to 5 min. Type II waves also have approximately 3-fold greater amplitudes than do Type I waves. Probably some of the waves called Type I are Type II, since only wave period is being measured; however, the greatest number of contractions are undoubtedly Type I.

increases, although the 10- and 14-hr. periods are probably indistinguishable.

Pain.—The subjective data for pain were combined to form a single matrix. The *Ss* were told to use whatever criteria they considered relevant in the light of their own experiences to define pain. The mean ranks for the 0-, 2-, 3-, and 4-min. immersions were 1.0, 2.7, 3.0 and, 3.2 respectively. The coefficient of concordance, .627, was significant beyond the .01 level. However, the intervals among the ranks are highly unequal, and in fact, the 2-, 3-, and 4-min. conditions were about equally rated. Thus there probably is little basis for concluding that the different exposure periods produced graded pain. There seems little doubt, however, that 2-, 3-, or 4-min. exposure to iced water does produce the sensation of pain, since all *Ss* agreed that these periods were painful.

It should be noted that there was a significant room temperature bias of the "order effect." Since the "cold-pressor effect" is insignificant and orthogonal to the "order effect," no corrections were applied. The mean room temperatures during the four cold-pressor sessions ranged from 24.1° C. to 24.4° C. Another interesting finding was that the data which required correction for skewness in the hunger analyses required the same corrections for pain. Thus rate of salivary flow and finger pulse volume changes were transformed to square roots, skin resistances to log conductances, and finger temperatures to log (37°-finger temperature).

Table 4 shows the effect of pain stimulation on the various autonomic functions and Table 5 the results of the *F* tests for those functions showing significant change. Systolic and diastolic blood pressures, finger pulse

TABLE 5

ANALYSES OF VARIANCE OF THOSE RESPONSES SHOWING SIGNIFICANT CHANGE DURING FOUR PERIODS OF EXPOSURE TO ICED WATER[a]

Source	*df*	Blood Pressure				Heart Rate		Finger Pulse Volume[c]	
		Systolic		Diastolic[b]					
		MS	*F*	*MS*	*F*	*MS*	*F*	*MS*	*F*
Cold Pressor	3	488.57	6.67***	533.35	11.22***	177.31	3.93*	.363	5.85**
Linear	1	1097.05	14.98***	724.50	15.24***	62.48	1.38	.220	3.55
Quadratic	1	368.16	5.03*	830.16	17.46***	73.53	1.63	.766	12.35**
Cubic	1	.49	—	45.38	—	395.60	8.77**	.105	1.69
Order	3	41.67	—	13.10	—	31.58	—	.033	—
Ss	15	238.98	3.26**	221.66	4.66***	179.00	3.97***	.461	7.43***
Error	42	73.24		47.54		45.09		.062	
Totai	63								

[a] See Footnote a, Table 3.
[b] Gap between the control and 3-min. exposure significant beyond the .05 level (gap = 12.5; R_p = 12.3 at the .05 level.
[c] Error *df* = 38, total *df* = 59.
* *P* = .05.
** *P* = .01.
*** *P* = .001.

volume changes and heart rate all had significant *F*'s, and there was a significant gap (Duncan, 1955) between the diastolic blood pressure for the control session and the 3-min. exposure. All those functions which yielded significant *F*'s also had significant trend components. Systolic and diastolic blood pressures had both linear and quadratic trends; finger pulse volume had a quadratic trend only; and heart rate showed a cubic component. There were 27 instances during which stimulation was preceded by some evidence of gastric motility. During 15 of these exposures, motility was completely blocked; in one *S* motility was induced by the stimulus on two occasions. The mean period for the 0-, 2-, 3-, and 4-min. conditions was 19.8, 19.4, 17.0, and 19.0, respectively; however, it should be noted that the stimulus means are less reliable than the prestimulus data; the *N* for the control condition was 12, whereas the 2-, 3-, and 4-min. means are based on *N*'s of 3, 5, and 6, respectively.

Discussion

Hunger.—Because of the parallel trends between the pulse pressure and systolic blood pressure curves, it seems probable that the pulse pressure effect is due essentially to changes in systolic pressure. It should be noted, however, that the diastolic pressure changes cannot be wholly ignored for these must have covaried with the systolic pressures thereby making pulse pressure changes significant. One of the most interesting findings is that many of the significant changes which occurred are attributable to the 10-hr. deprivation period: pulse pressure was least, systolic pressure lowest, and salivary flow slowest under this condition. In the case of the blood pressure changes, the 24-hr. deprivation effects were indistinguishable from those of the control period. The rate of salivary flow, however, increased fairly regularly with increased deprivation, after the initial drop. Axillary temperature, respiration rate, and the period of Type 1 stomach contractions all showed fairly linear changes to the deprivation periods. The first manifested a gradual increase as a function of deprivation time, while the last two showed progressive decreases.

The finding that axillary temperature increases with increased food deprivation has a bearing on Strominger and Brobeck's (1953) thermodynamic theory. It would seem that they would predict just the opposite effect. However, it should be remembered that while axillary temperature showed a statistically significant linear trend, the differences among the means for the four conditions were not significant. On the other hand, Booth and Strang's (1936) data, which Strominger and Brobeck cite as support for their position, were based on temperature levels of the hands and feet. These probably more accurately reflected blood flow under the sensing element than body temperature per se. In the present study there were no differences between face and finger temperatures during the control tests and the various food deprivation conditions. A critical test of Strominger and Brobeck's hypothesis would result from measurement of temperature changes in the hypothalamus itself.

The breathing rate effects are consistent with Cannon's (1929) hypothesis that increased hunger results in heightened activity of the cranial division of the ANS. Likewise, the 10-hr. drops in systolic blood pressure and the 14- and 24-hr. increases in rate of salivary flow could be attributed to parasympathetic (PNS) hyperactivity; and the increase in incidence of Type 1 stomach waves is indicative of heightened vagal activity. But certain of the data do not support the thesis that hunger is univocally associated with heightened activity of the cranial ANS. The 10-hr. decrease in salivary flow, the reversal in trend of systolic blood pressure, the progressive increase in body temperature, and the

decreased period of gastric contractions could possibly be attributed to heightened SNS activity. These data point to two conclusions: (*a*) that the cranial ANS is not the only factor operating to alter peripheral autonomic responses during hunger; (*b*) that the relationship between the intensity of hunger and level of activation of specific autonomic variables is not linear.

In order to account for the sympathetic-like effects in the presence of some indices of increased PNS activity, it seems necessary to infer some increase in level of activity of (*a*) portions of the SNS, (*b*) the adrenal medullae, or both. Which factors are actually operant must be the subject for future research.

The subjective data point to the conclusion that there was, in general, an increase in hunger with increased deprivation time. If increases in intensity of hunger produce concomitant increases or decreases in the level of activation in any branch of the autonomic nervous system, one should expect monotonic trends in all functions mediated by that branch. Instead, it was observed that the pressor responses had quadratic rather than linear trend components. It therefore follows that the pattern as well as the level of autonomic response differs for different periods of food deprivation.

Pain.—The major effects found under cold-pressor stimulation were cardiovascular changes. Both systolic and diastolic pressures increased about equally, thus leaving pulse pressure unaffected; and there were increases in heart rate and peripheral resistance. The fact that the stimuli blocked gastric contractions in over 50% of the cases and induced it twice are also pertinent results.

The finding that the cold-pressor test will produce a significant pressor response is, of course, not new; nor is it surprising that this effect is mediated through vasoconstrictor and cardio-accelerator reflexes. More interesting are some of the statistically insignificant results: (*a*) that skin conductance level is not a function of painful stimulation; (*b*) that increased breathing rate is not a necessary consequence of painful stimulation; and (*c*) that there is not a significant drop in finger temperature with such stimulation.

There are several factors that could account for these apparent discrepancies. One is that the quality of stimulation determines the nature of the response. Thus there might be different patterns of response associated with different pain-producing stimuli. A second possibility is that the effects others have attributed to painful stimuli may be associated with startle at the onset of the stimulus rather than with the stimulus per se. If this is true, there may be a generalized response to the onset of stimulation, out of which develops more specific response patterns to specific classes of stimulation. It should be pointed out that the measure of skin conductance used here was the level near the end of the stimulus period (cf. Table 1) rather than the maximum change. However, in a number of instances there was no change in skin resistance at any time during the stimulation period.

In view of the dissimilarities in trend among the various responses, it appears that different durations of cold-pressor stimulation resulted in different patterns of autonomic activation despite the fact that the subjects did not clearly distinguish differences in pain among the three durations of exposure. One possible explanation of the apparently dissimilar patterns is that some adaptation was taking place. Table 4 shows this effect occurred in the pressor responses and in finger pulse volume changes. Heart rates also tended to decrease during the longer exposure periods.

Summary

This study reports the effects of 2, 10, 14, and 24 hr. of food deprivation (hunger) and of 0, 2, 3, and 4 min. of cold-pressor stimulation (pain) on 12 autonomically mediated responses of 16 male *Ss*. It was observed that increased food deprivation led to: (*a*) decreases in pulse pressure which probably were primarily a result of decreases in systolic blood pressure; (*b*) linear increases in axillary temperature and salivary output; (*c*) a linear

decrease in respiration rate; (*d*) increased incidence of Type 1 stomach contractions; and (*e*) decreased period of Type 1 contractions.

Cold-pressor stimulation resulted in: (*a*) increases in systolic and diastolic blood pressures; (*b*) increases in heart rate; (*c*) decreases in finger pulse volume; (*d*) some blocking of Type 1 stomach contractions (56% of the cases); or (*e*) decreased period of Type 1 contractions.

It was concluded that: (*a*) the thermodynamic theory of hunger was not supported; (*b*) the relationship between intensity of hunger and level of autonomic response is not linear; (*c*) the pattern of response to immersion in iced water varied with the duration of immersion even though *Ss* did not report any differences in pain among the immersions.

References

Anand, B. K., Dua, S., & Shoenberg, K. Hypothalamic control of food intake in cats and monkeys. *J. Physiol.*, 1955, **127**, 143–152.

Booth, G., & Strang, J. M. Changes in temperature of the skin following the ingestion of food. *Arch. int. Med.*, 1936, **57**, 533–543.

Brobeck, J. R., Tepperman, J., & Long, C. N. H. Experimental hypothalamic hyperphagia in the albino rat. *Yale J. Biol. Med.*, 1942, **15**, 831–853.

Cannon, W. B. *Bodily changes in pain, hunger, fear, and rage.* New York: Appleton-Century, 1929.

Code, C. F., Hightower, N. C., Jr., & Morlock, C. C. Motility of the alimentary canal in man: Review of recent studies. *Amer. J. Med.*, 1952, **13**, 328–351.

Duncan, D. B. Multiple range and multiple F tests. *Biometrics*, 1955, **2**, 1–42.

Strominger, J. L., & Brobeck, J. R. A mechanism of regulation of food intake. *Yale J. Biol. Med.*, 1953, **25**, 383–390.

Wenger, M. A. Studies of autonomic balance in Army Air Forces personnel. *Comp. Psychol. Monogr.*, 1948, **19**, 1–111.

Wenger, M. A., Engel, B. T., & Clemens, T. L. Studies of autonomic response patterns: Rationale and methods. *Behav. Sci.*, 1957, **2**, 216–221.

Wenger, M. A., Henderson, E. B., & Dinning, J. S. Magnetometer method for recording gastric motility. *Science*, 1957, **125**, 990–991.

Wolf, S., & Hardy, J. D. Studies on pain. Observations due to local cooling and on factors involved in the "cold-pressor" effect. *J. clin. Invest.*, 1941, **20**, 521–533.

Wolff, H. G., & Wolf, S. *Pain.* Springfield, Ill.: Thomas, 1951.

(Received June 25, 1958)

Journal of Experimental Psychology
Vol. 57, No. 6, 1959

THE CONCURRENT EFFECTS OF PROACTIVE AND RETROACTIVE INHIBITION

ROBERT J. SEIDEL [1]

University of Pennsylvania

In the past (McGeoch & Irion, 1952) retroactive interference (RI) had been considered the major contributor to forgetting. A recent review of the literature, however, had led to the implication that one principal influence heretofore ignored is actually proactive interference (Underwood, 1957). The function obtained by Underwood through his collation of the data from classical studies on forgetting indicates that approximately 85% of the effect, when corrected for artifactual recall limit, *can be* directly attributable to the number of previous lists (PI) given. Although not specifically considered by Underwood, the relative importance of such a function for the laboratory study of forgetting, however, rests in its relation to a comparable RI curve and a combined RI and PI curve. Since the above noted results were obtained from studies having no formal RI material, the question remains of determining within the formal experimental setting, the relative contributions of both sources of interference. Only one previous classical study (Lester, 1932) presents any indication of a possible summative effect of RI and PI. However, the result was inadvertent to the main purpose of the study, and no systematic investigation of combinative RI and PI effects was attempted. It is this problem of concurrent effects which the present study investigated.

[1] Now at Human Factors Division, Research Directorate, Kirtland Air Force Base, New Mexico.

METHOD

Subjects.—A sample of 90 *Ss* was used in a 3 × 3 factorial design. The *Ss* were mostly undergraduate summer school students taking introductory psychology. The remainder of the sample included some students of an advanced psychology course as well as a few departmental secretaries. The students in introductory psychology were given credit for a laboratory report, but the other *Ss* volunteered without reward.

Apparatus.—A 35-mm. Revere automatic slide projector was used to present nonsense syllables at a rate of 4 sec. per syllable. There was a 2-sec. intersyllable interval. The nonsense syllables were drawn from the 53%, 47%, and 40% Glaze lists (Hilgard, 1951), and were divided into nine lists each of 12 syllables with equivalent (47% median) associative value. The lists were assigned at random to the cells for use as first lists. Because of the possible effects of rehearsal among classmates, nine lists (rather than the five required for counterbalancing) were used to discourage communication among *Ss*. Each list was then used a total of three times, once as OL, PL, IL with each cell having a different list for OL and a different combination of lists for PL and IL. Hilgard's (1951, p. 540) suggested rules for use of nonsense material were followed.

Procedure.—The basic paradigm for this study was X-A-B-A. That is, in Session 1 *Ss* were given either zero (P_0), one (P_1), or two (P_2) practice lists (X). In Session 2 they were given a test list (A) and in Session 3 they were given zero (R_0), one (R_1), or two (R_2) RI lists (B) immediately prior to the recall of A. Twenty-four hours separated each session. During the planning of the experimental design, the question of appropriate intertreatment time intervals was considered in detail. The IL Recall interval was based upon available literature (Archer & Underwood, 1951; Briggs, 1954) which had shown no change in RI effects over 48 hr. However, the data on the effects of varying PI Recall temporal interval had not yielded clearcut results prior to Underwood's recent studies (1948; 1949). When considered

together, these two experiments reveal that the time interval chosen in the present study, 48-hr., maximizes the possible influence of PI on recall. Moreover, other evidence (Melton & Von Lackum, 1941; Ray, 1945) indicates the maximal effect of PI on transfer to be between 24–48 hr., which tends to justify the 24-hr. PI-OL interval used.

The experiment was run as a two-factor variance design with 10 *Ss* assigned at random to each cell. By means of the method of anticipation each list was learned to one errorless trial. The *Ss* of each cell were run in a group, thereby yielding nine groups of 10 each. Broken appointments necessitated filling the cells with the following additional *Ss*, also run in groups where possible: P_0R_0, 1; P_0R_2, 1; P_1R_0, 5; P_1R_1, 4; P_2R_0, 1; P_2R_1, 2. Anticipations were written on a spiral pad and after each trial, *S* turned to a new page. As each *S* reached criterion, he notified *E* and then proceeded to an irrelevant task inserted to obscure the actual purpose of the experiment. The *E* continually checked the writing of the anticipations during the experiment, and the group technique of collecting the data proved extremely effective. Only one *S*'s data were not counted because she wrote down the syllables after they were exposed. Both recall and learning were measured. Recall score was the number of correct anticipations on the first relearning trial. Relearning score was number of trials to one errorless repetition. Responses were not forced during recall and relearning.

Although the purpose of the experiment was to study forgetting as function of interference, *Ss* were told that the experiment was an attempt "to study the effects of verbal learning on motor problem solving." A jigsaw puzzle was utilized to hide the actual purpose. It was hoped that with this subterfuge, motivation to learn, social pressure, and rehearsal between sessions would all be properly controlled. Prior to the PI session (X), P_1 and P_2 *Ss* were instructed that before starting "the main portion of the experiment tomorrow, I am going to give you practice in the tasks in order to get you acquainted with the experimental materials." To lessen possible tension, stemming from the group setting, assurance was given that many *Ss* were being used simultaneously to facilitate collecting data and that *Ss* should pay no attention to their neighbors' work. Code numbers were used to identify the work of each *S*. As *Ss* reached criterion on the list, they were given difficult jigsaw puzzles to solve. Each *S* kept his own record of time and number of pieces correctly joined together.

As soon as the last person in the group reached criterion on the list, puzzle solving was stopped, and dependent on the experimental condition, the group learned a second list or were dismissed for the day. At this point those who had not had time to work on the puzzles were reassured that they would have "an opportunity to do so before the experiment is completed." At the end of each session, *Ss* were asked to forget how their performance seemed that day, to relax between sessions, and to be ready to start fresh on the following day.

That the instructions and the puzzle were effective was clearly shown at the end of the experiment. First, when *Ss* were told after learning B that the next list they were to learn was the same list which they had been given the previous day (OL) all expressed surprise and bemoaned such an undertaking. Secondly, in a postexperiment interview, no student guessed the purpose of the experiment and nearly all had accepted the stated purpose without question. Finally, no student reported rehearsal of lists between sessions.

Results

Although specific first lists differed among all groups, an analysis of variance of learning first lists showed no significant differences existed initially between the nine groups ($F = 1.37$, 8 and 81 *df*). Furthermore, a similar analysis on the learning of the test list (OL), also different for each cell, revealed no differential practice effects based upon different amounts of PI learning ($F = 1.42$, 8 and 81 *df*).

Table 1 presents the mean percentages recalled and trials to relearn for each treatment group. The Bartlett test of homogeneity of variance computed on both the recall and relearning data was not significant in either case. Both main effects were significant for recall beyond the .001 level (see Table 2). There was no significant interaction. The relearning data essentially substantiate these findings although a significant interaction appears in Table 2. The only PI simple effect which is significant occurs at the R_2 level ($F = 5.16$,

$P < .01$). The RI effects are significant at the P_1 and P_2 levels ($F = 3.7$, $F = 7.59$, $P < .01$). Critical differences between cell means indicate three significantly different groupings in recall: the zero interference condition, the pure RI and PI conditions, and the mixed interference conditions. In relearning the difference between no interference and the pure PI and RI conditions drops out.

Further analysis, based upon the quality of interference, yielded effects not apparent in the critical-t comparison. Considering as treatments the zero interference condition, pure PI, pure RI, and the combined PI and RI conditions, the data were next analyzed as a simple randomized design of four groups. The dimension consisted of type of interference. The F ratios for both recall ($F = 57.5$, 3 and 86 df) and relearning ($F = 8.7$) were significant ($P < .001$). Figure 1 shows clearly four different levels of recall dependent upon the nature of the interpolated activity. None of the intralevel comparisons were significant, a finding which indicates that the variations from one list to two, etc., at any given interference level were random. The reason for this is probably that the 12-syllable nonsense list of 47% associative value learned by anticipation was not a sufficiently sensitive task. Interlevel comparisons were all significant, however, beyond the .01 level. For the smallest difference (between pure PI and pure RI) $t = 4.6$ and $P < .01$ ($df = 38$). Supplementary information on the type of interference obtained is given in Table 1. The interlist intrusions at recall clearly show a predominance of proactive words (an effect which can be expected with the relative temporal conditions utilized).

TABLE 1

MEANS FOR PERCENTAGE RECALL, FOR RELEARNING TRIALS, AND FREQUENCY OF INTRUSION DURING RECALL OF OL FROM PROACTIVE AND RETROACTIVE SOURCES

Cond.	Percentage Recall		Trials to Return		Intrusions	
	Mean	*SD*	Mean	*SD*	P	R
P_0R_0	72	26.9	3.3	1.64	—	—
P_1R_0	46	17.2	3.7	.82	5	—
P_2R_0	48	17.2	3.2	.6	6	—
P_0R_1	34	24.6	4.2	1.62	—	1
P_0R_2	40	23.2	3.2	1.48	—	2
P_1R_1	9	11.4	5.3	2.00	4	1
P_2R_1	16	15.4	5.5	1.44	7	0
P_1R_2	7	11	4.9	1.1	4	1
P_2R_2	10	14.1	4.9	1.1	5	1

Note.—Critical difference for .05 level of confidence: 17.85% Recall, 1.22 trials Relearning.

TABLE 2

ANALYSES OF VARIANCE ON RECALL AND RELEARNING SCORES

Source	*df*	Recall			Relearning		
		MS	*F*	*P*	*MS*	*F*	*P*
PI	2	104.03	20.6	<.001	4.8	2.57	NS
RI	2	180.6	35.7	<.001	13.8	7.38	<.01
PI × RI	4	1.97	.39	NS	5.25	2.80	<.05
w	81	5.05			1.87		

DISCUSSION

In terms of comparison with Underwood's review (1957), it is most important to note the 72% recall for the zero interference condition (P_0R_0) in the present investigation. This corresponds almost identically with the zero point of Underwood's PI curve and thus adds strength to the comparability of the two sets of data. In a larger sense, this agreement means that Underwood's general conclusion that experimental manipulation accounts for 75% of the usual obtained variance in forgetting is verified. (Note that this stability of occurrence appears despite variations in apparatus, tasks, etc.) In addition, as seen in Table 1 of the present study, the effect of two proactive lists, 48%

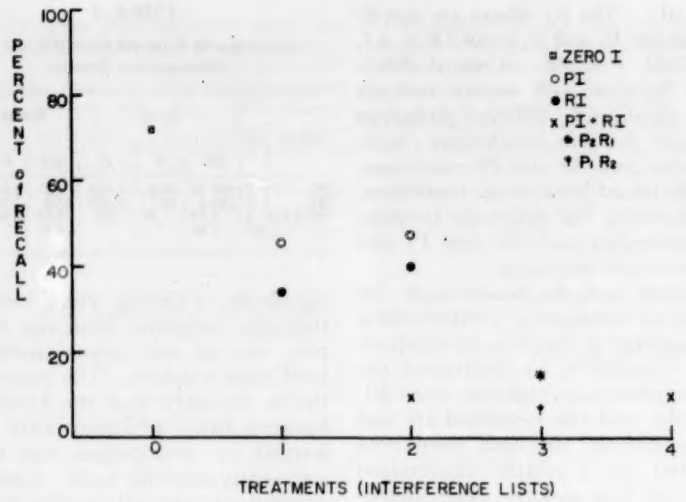

FIG. 1. Recall and type of interference.

recall, was almost identical to the recall (approximately 47% for two warm-up lists) shown in Underwood's curve.

It is most interesting to note that the present and the collated data yield similar upper limits. However, as seen in Fig. 1 the slope of any given recall function is dependent on the quality of the interfering material. For example, given the temporal intervals in the present study, the curve of the combined interference data is much steeper than either PI or RI alone. Further, one previous retroaction study comparable in part to the present experiment presents similar data. Lester's (1932) experiment included a group given one retroactive list in addition to the two practice lists given everyone (PI); and for the RI Group recall was 3.3% as compared with 35.3% for the corresponding PI Group. The comparable cells in Table 1 show a similar drop from 48% recall to 16% with the inclusion of one RI list. Going from two to three lists in Underwood's curve, however, results in a slight drop from 47% to approximately 42% (Underwood, 1957, p. 53). What the latter curve represents, of course, is a forgetting function based solely on PI. With as many as 21 PI lists, the monotonic function reaches almost total forgetting as a limit. Certainly increasing PI or RI may result eventually in complete experimental forgetting. The value of the present study, however, rests in showing, given the comparability to Underwood's collated data, that the combined PI-RI interference function arrives at this point much more rapidly. Furthermore, in view of the practical inevitability for the occurrence of at least some informal PI and RI in any laboratory experiment on forgetting, a paradigm containing both interference components seems to be the realistic one for the experimental study of forgetting.

Incidental to the main purpose of the investigation discussed above, the present study also has produced data relevant to the process of forgetting. Although the data are not crucial to the issue, the intrusions do, as shown in Table 1, provide some interesting information concerning a one- or two-factor theory (interference and unlearning) of forgetting (Melton & Irwin, 1940). The

theoretical argument goes that, if competition alone is responsible for forgetting, then the learning of interpolated lists should be interfered with to the same degree by prior learning (PI) as the recall and relearning of the original list is by interpolated learning (RI). In the most general case, the proponents of two-factor theory have stated that in order for a competing response or interference theory to be correct, RI should equal PI (Underwood, 1945). Note that Fig. 1 shows pure PI to yield greater recall than pure RI. It is true that such an effect has previously been reported; hence, the two-factor theory. However, in none of these classical studies on forgetting (McGeoch & Irion, 1952; McGeoch & Underwood, 1943; Melton & Irwin, 1940; Melton & Von Lackum, 1941; Underwood, 1945) were PI and RI investigated concurrently. Typically one would measure PI or RI with one group of subjects or set of lists and then measure these effects with the other separately. The difference obtained between the curves was said to be due to unlearning. The present study at this point seems to support this view.

Indication that interference and unlearning may not operate as customarily conceived, however, is found in Table 1 by the predominance of proactive intrusions at recall.[2] If unlearning influences forgetting, then it follows that material which has more similar response learning subsequent to its being learned should be subject to more unlearning and thus to greater forgetting than material having less such learning following. This means that the proactive lists in the present study had a greater probability of being influenced by unlearning than did the retroactive lists. It also follows that, if greater unlearning occurs for PI, then it should exert less overt interference (e.g., intrusions) on recall than the RI. Yet the data in Table 1 show the intrusion frequency to be reversed, and thus are contrary to the unlearning hypothesis. From a strict interference viewpoint, however, one can readily account for this outcome. To begin with, overt interference (intrusions) is greater when the competing response systems are less discriminable. Response systems in turn are differentiable in inverse relation to time between their learning and the recall of one system, e.g., the test list (Underwood, 1945). Note that strong interference or overt intrusions indicates poor response differentiation. Given the procedure in the present study, this interpretation would predict more overt PI (lists learned 48 hr. prior to recall) and greater implicit (weaker) RI (lists learned immediately prior to recall). Table 1 agrees with the prediction.

Although the present analysis does indicate, then, general agreement with empirical findings on interference and supports a general interference theory of forgetting, the data do oppose a simple PI or RI formulation. Actually, in terms of a competing response interpretation which includes both PI and RI the results make good sense. When either proactive lists or retroactive lists are introduced alone, they act to compete unidirectionally with the responses of the test list. But when the interfering materials are combined, the test list is acted upon from two directions at recall, prior learning, and interpolated learning. Of course, the question of the reasons for the difference in related strengths of RI and PI still remains. To speculate a bit further, one might consider the learning of many lists simply serial learning on a larger scale. In any serial list the words between similar words are learned with greater difficulty than those benefiting from primacy or recency. Similarly, recalling the first or last list learned in a series may be easier than recalling lists learned in the middle of the series. In fact, this interpretation suggests the possibility of an interlist serial recall curve analogous

[2] An additional difficulty for any simple interpretation is that, at least in one case (Underwood, 1948), the difference between RI and PI was found to disappear as the OL-Recall interval was increased from 5 to 48 hr. Although RI did not change, PI increased. In this study, PI lists occurred immediately prior to OL.

to the usual intralist serial learning bow. If the analogy holds, it would mean that other serial phenomena such as the von Restorff effect should also be found in interlist recall studies.

Summary

This experiment was designed to investigate concurrent effects of PI and RI on forgetting. Ninety *Ss*, 10 per cell, were used in a two-factor design consisting of zero, one, or two amounts of PI and/or RI. Memorization of lists of nonsense syllables was used as the experimental tasks. Recall and relearning were measured.

The data supported a general interference theory: (*a*) Increasing number of interfering lists learned prior to recall of the test list, and (*b*) quality of interfering lists, viz., pure PI, pure RI, or combined PI-RI, yielded increasing amounts of forgetting.

References

Archer, E. J., & Underwood, B. J. Retroactive inhibition of verbal associations as a multiple function of temporal point of interpolation and degree of interpolated learning. *J. exp. Psychol.*, 1951, **42**, 283–290.

Briggs, G. E. Acquisition, extinction, and recovery functions in retroactive inhibition. *J. exp. Psychol.*, 1954, **47**, 285–293.

Hilgard, E. R. Methods and procedures in the study of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951.

Lester, O. P. Mental set in relation to retroactive inhibition. *J. exp. Psychol.*, 1932, **15**, 681–699.

McGeoch, J. A., & Irion, A. L. *The psychology of human learning*. New York: Longmans, Green, 1952.

McGeoch, J. A., & Underwood, B. J. Tests of the two-factor theory of retroactive inhibition. *J. exp. Psychol.*, 1943, **32**, 1–16.

Melton, A. W., & Irwin, J. McQ. The influence of degree of interpolated learning on retroactive inhibition and the overt transfer of specific responses. *Amer. J. Psychol.*, 1940, **53**, 173–203.

Melton, A. W., & Von Lackum, W. J. Retroactive and proactive inhibition in retention: Evidence for a two-factor theory of retroactive inhibition. *Amer. J. Psychol.*, 1941, **54**, 157–173.

Ray, W. S. Proactive inhibition: A function of time interval. *Amer. J. Psychol.*, 1945, **58**, 519–529.

Underwood, B. J. The effect of successive interpolations on retroactive and proactive inhibition. *Psychol. Monogr.*, 1945, **59**, No. 3 (Whole No. 273).

Underwood, B. J. Retroactive and proactive inhibition after 5 and 48 hours. *J. exp. Psychol.*, 1948, **38**, 29–38.

Underwood, B. J. Proactive inhibition as a function of time and degree of prior learning. *J. exp. Psychol.*, 1949, **39**, 24–34.

Underwood, B. J. Interference and forgetting. *Psychol. Rev.*, 1957, **64**, 49–60.

(Received June 30, 1958)

Journal of Experimental Psychology
Vol. 57, No. 6, 1959

VARIABLES AFFECTNG CHILDREN'S PERFORMANCE IN A PROBABILITY LEARNING TASK[1]

HAROLD W. STEVENSON AND MORTON W. WEIR

University of Texas

It was found in a recent study (Stevenson & Zigler, 1958) that children's performance in a probability learning task differed for normal and institutionalized feebleminded children and for normal children given pretraining on games involving high and low degrees of success. It was hypothesized that these differences were related to the children's different expectancies for frequency of reinforcement. These expectancies were assumed to affect performance by determining the degree to which the child would accept a solution in the probability learning task which yielded less than 100% reinforcement. It was predicted that *S*s with low expectancies would show a greater frequency of choice of the reinforcing stimulus than would *S*s with higher expectancies. This would result from the attempt by *S*s with higher expectancies to seek, through variable behavior, a means by which they could obtain a frequency of reinforcement corresponding to the frequency which they expect.

In the present study three experiments are reported investigating the effects of three variables on probability learning in children: CA, incentive conditions, and shifts in percentage of reinforcement. It is assumed that older children have expectancies for a higher percentage of reinforcement than do younger children and that the performance of these groups will differ in the manner outlined above. In the experiment involving incentive conditions, the same general hypothesis is tested; however, it is assumed that the degree to which the child will accept a particular frequency of reinforcement is a function in part of the value of the reinforcement to *S*. In the third experiment, the effects of shifts in percentage of reinforcement during the experimental trials are investigated. The hypotheses tested in each experiment are discussed in greater detail in later sections.

[1] This study was supported by a grant from the National Science Foundation (Grant 3280) to the first author.

EXPERIMENT I

The purpose of this experiment was to investigate the behavior of young children of different chronological ages in a probability learning task. There is some basis for assuming that performance in such a task would differ as a function of CA. It has been suggested (Goodnow, 1955) that one of the conditions influencing performance in a probability learning task is the level of success *S* will accept in the task. It seems reasonable to assume that older *S*s' greater familiarity with soluble problems and their greater desire to master adult-controlled tasks would result in their being less likely than younger children to accept low levels of success. It is assumed that when less than 100% reinforcement is available older children would seek a solution which would provide a level of success more acceptable than that attainable by the consistent choice of the reinforcing stimulus.

The variable behavior resulting from their attempts would reduce the frequency of choice of the reinforcing stimulus below that found with younger *Ss*. It is predicted, therefore, that with levels of reinforcement below 100% there is an inverse relationship between frequency of choice of the reinforcing stimulus and chronological age. It is further predicted that under conditions of 100% reinforcement the asymptotic frequency of correct response will be comparable at all CA levels, since 100% reinforcement provides a level of success acceptable to all *Ss*.

Method

Subjects.—The *Ss* were 120 children attending nursery and elementary schools in Austin, Texas.[1] The *Ss* were selected at random from among children of the appropriate CA's enrolled in the schools. Children at four age levels were used: 3–0 to 3–11, 5–0 to 5–11, 7–0 to 7–11, and 8–9 to 10–2. There were 30 *Ss* at each CA level.

Apparatus.—The apparatus has been described previously (Stevenson & Zigler, 1958). It consisted of a yellow panel containing a horizontal row of three knobs. Above the knobs was a signal light and below the knobs was a delivery hole for marbles. The marbles fell into an enclosed plastic container.

Procedure.—The *S* was seated in front of the apparatus and was told that he was to play a game. The *E* demonstrated the apparatus, and said, "When the light comes on, you push one of the knobs. If you push the correct knob a marble comes out here, like this. Now every time the light comes on you push the knob that you think will get you the marble. Remember, just push one knob each time the light comes on." The *S* was told that he was to get as many marbles as he could and when the game was over he could choose two toys from a selection of toys including balloons, plastic figures, etc., which *E* showed him.

Three conditions which differed in the percentage of reinforcement of correct response were used. For each *S*, one of the three knobs (either L, M, or R) was designated as the correct knob. The knob which was to be correct for a particular *S* was determined by a prearranged random sequence. An equal number of *Ss* was tested with each knob. The correct knob yielded reinforcement; choices of the other two knobs were never reinforced. Depending upon the condition, the correct knob yielded 100%, 66%, or 33% reinforcement. In other words, a contingent procedure of reinforcement was employed whereby *S* was reinforced for a certain percentage of the times that he chose the correct knob. For example, in the 33% condition reinforcement was delivered on 33% of *S*'s choices of the correct stimulus rather than on 33% of the trials. Ten *Ss* at each CA level were assigned at random to each of the three conditions. In the 66% and 33% conditions the trials on which a choice of the correct knob was reinforced were determined by a random sequence of 30 trials. This sequence was repeated when *S* pressed the correct knob on more than 30 trials. The *Ss* were given a total of 80 trials each.

Results and Discussion

The average numbers of correct responses in blocks of 10 trials for the four age levels and the three percentages of reinforcement are presented in Fig. 1. An examination of the curves for the 33% and 66% groups reveals that performance differed as a function of CA. As predicted, there is a tendency for older *Ss* to choose the reinforcing stimulus less frequently than younger *Ss*. An analysis of variance of the results of the total 80 trials indicates a significant difference associated with CA ($F = 4.43$, $df = 3$ and 72, $P < .01$). A significant difference is found also for percentage of reinforcement ($F = 7.37$, $df = 1$ and 72, $P < .01$). The inter-

[1] The writers wish to thank Larry Schmucker for testing *Ss* in Exp. I, Mary Bess Whidden for testing *Ss* in Exp. II, and Martin Steigman for testing *Ss* in Exp. III, and the directors or principals of All Saints Nursery School, Austin High School Child Development Laboratory, University of Texas Nursery School, Tarrytown Methodist Nursery School, and St. Louis School, St. Austin School, St. Mary's School, and Trinity Lutheran School for their cooperation in making *Ss* available.

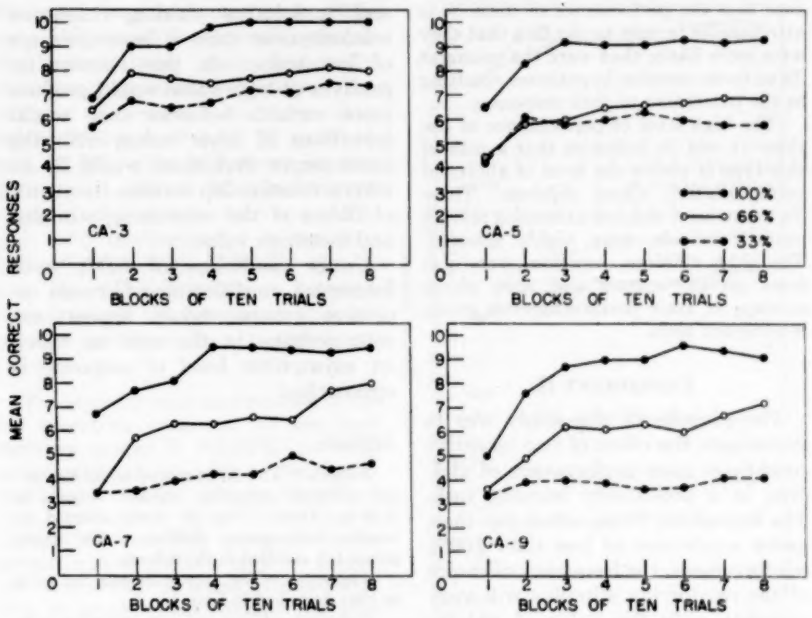

FIG. 1. Mean number of correct responses for blocks of 10 trials: Exp. I.

action between CA and percentage of reinforcement is not significant ($F<1$).

Performance in the groups receiving 100% reinforcement was similar at each age level. It was predicted that asymptotic level of response would be comparable in all four groups. This was found to be true, for all but six *Ss* chose the correct stimulus on all of the last 20 trials. There were two *Ss* in each age group except CA 3 who did not consistently choose the correct knob during the last 20 trials.

There are several additional features of *Ss'* performance which are of interest. The *Ss* initially had strong position preferences. On Trial 1, 15% of the *Ss* chose the left knob, 50% the middle knob, and 35% the right knob. As a control for possible knob preferences, each knob was correct for an equal number of *Ss* at each CA level. Knob preferences did not operate throughout all the trials, however, for of the total number of correct responses made on the 80 trials, .33 were made when the left knob was correct, .35 when the middle knob was correct, and .32 when the right knob was correct.

In order to determine whether the poor performance of older *Ss* was a result of their developing patterns of response, each individual's performance was examined. A high frequency of LMR and RML patterns of response appeared. The *Ss* at CA 3 had an average of 8.4 repetitions of these patterns, *Ss* at CA 5 an average of 14.8, *Ss* at CA 7 an average of 14.5, and *Ss* at CA 9 an average of 14.0. No other patterns of response were readily apparent. The behavior of *Ss* at CA's 5, 7, and 9, therefore, was not random since approximately 50% of the responses were made in the form of one of the patterns analyzed. This may indi-

cate that the performance of these *Ss* is attributable in part to the fact that they were more likely than were the youngest *Ss* to form complex hypotheses resulting in the patterning of their responses.

The high level of performance of the three-yr.-old *Ss* indicates that a task of this type is within the level of ability of young nursery school children. These *Ss*, as are most children attending private nursery schools, were highly selected. The older children, however, were also from private schools and were above average in their performance on group intelligence tests.

Experiment II

The purpose of this study was to investigate the effect of two incentive conditions upon performance of children in a probability learning task. The hypothesis to be tested was that, under conditions of less than 100% reinforcement, the frequency of choice of the reinforcing stimulus will vary inversely with the value of the incentive. The hypothesis is derived in the following manner. In simple probability learning tasks such as the one used, children should learn rather quickly that the reinforcing stimulus does not pay off 100% of the time. If *S* accepts this frequency of reinforcement, *S* should choose the reinforcing stimulus with a high degree of consistency. If this frequency of reinforcement is not acceptable, *S* must attempt to find a solution to increase the frequency with which he is rewarded. This attempt leads to variable behavior, with the result that the frequency of choice of the reinforcing stimulus is lower than that occurring if such an attempt is not made. It is assumed that incentives affect performance by determining the degree to which *S* would attempt to find such solutions. It is further assumed that if incentives are of high value *S* is more likely to seek a solution yielding consistent reinforcement than if incentives are of low value. In this manner incentives of high value would produce more variable behavior than would incentives of lower value, with the consequence that there would be an inverse relationship between frequency of choice of the reinforcing stimulus and incentive value.

Under conditions of 100% reinforcement no differences between incentive groups should appear, except perhaps in the rate at which an asymptotic level of response is approached.

Method

Subjects.—The *Ss* consisted of 60 five-yr.-old children attending nursery schools in Austin, Texas. The *Ss* were selected at random from among children of the appropriate CA enrolled in the schools.

Apparatus.—The apparatus was the same as that described in Exp. I.

Procedure.—High and low incentive conditions were used. In the high incentive condition, the choice of the reinforcing knob resulted in the delivery of a trinket. A wide assortment of small plastic trinkets was provided. In the low incentive condition, the choice of the reinforcing knob resulted in the delivery of a marble. The *S* was told that at the end of the experiment the marbles could be exchanged for three trinkets of the types used in the high incentive condition.

As in Exp. I, reinforcement was delivered on 33%, 66%, or 100% of the choices of the correct knob. The procedure was the same as that in Exp. I except for the reinforcements used.

Results

The average numbers of times that the correct knob was selected in each incentive condition under the three percentages of reinforcement are presented in Fig. 2. The curves for the 100% reinforcement groups are highly similar and do not differ significantly ($t = .12$). The data in Fig. 2 support the prediction that

the low incentive groups in the 66% and 33% conditions would choose the correct knob more frequently than the high incentive groups. An analysis of variance of these data results in an F of 6.71 ($P = .05$, 1 and 36 df). The difference in frequency of choice of the correct knob in the groups receiving 66% and 33% reinforcement is not significant ($F = 2.70$, $P > .05$, $df = 1$ and 36). The interaction between percentage of reinforcement and incentive condition is not significant ($F < 1$).

The higher level of performance in the low incentive, compared to the high incentive groups in the 33% and 66% conditions is an inversion of the results typically found in learning situations. It has been noted earlier (Stevenson & Zigler, 1958) that even though *Ss* in the present type of problem may not choose the reinforcing stimulus consistently, they do know which stimulus has provided reinforcement. This suggests that performance in this situation does not mirror the learning that has taken place; rather, it indicates the degree to which *S* will accept the percentage of reinforcement provided by the reinforcing stimulus.

The two incentive conditions appeared to operate in the anticipated manner; *Ss* in the high incentive groups seemed to be much more pleased to get the trinkets than were *Ss* in the low incentive groups to get the marbles.

These results differ from those of Messick and Solley (1957), who report exploratory data indicating an increase in frequency of choice of the reinforced stimulus when candy was compared to verbal reinforcement as the incentive. It is difficult to compare the results of the two studies because of differences in procedure and because of the descriptive nature of their results.

The results may be interpreted as providing support for the suggestion by Goodnow (1955) that one of the conditions affecting the distribution of

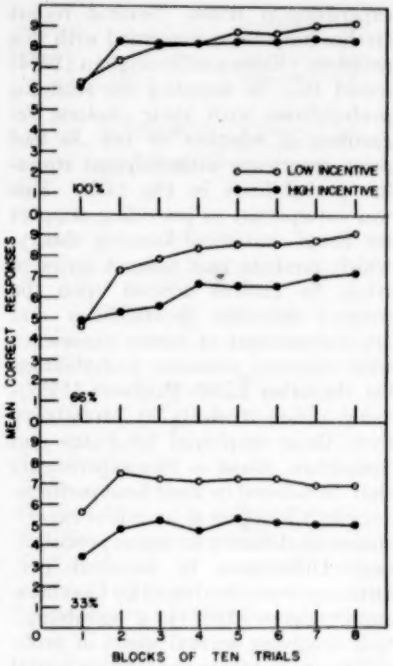

FIG. 2. Mean number of correct responses for blocks of 10 trials: Exp. II.

choices in a probability learning task is *S*'s ability to overlook losses. If the reward has significance for *S*, he may be less likely to overlook loss than when the reward has less significance. The *S*'s inability to overlook losses, as Goodnow suggests, would lead to more variable behavior than when *S* is less concerned about the failure of a response to lead to reinforcement. Behavior would become more variable as *S* attempts to seek a solution to minimize the frequency of nonreinforced responses.

EXPERIMENT III

The purpose of this experiment was to investigate the performance of *Ss* in a probability learning task when the probability of reinforcement is changed during the course of the

experimental trials. Several recent studies have been concerned with this problem. Estes and Straughan (1954) found that *Ss* matched the stimulus probabilities with their choices, regardless of whether or not *Ss* had prior experience with different stimulus probabilities in the task. This was interpreted as providing support for Estes' statistical learning theory, which predicts that after a series of trials *Ss*' choices depend upon the present stimulus probabilities and are independent of earlier experience with different stimulus probabilities. On the other hand, Parducci (1957), using different shifts in probability from those employed by Estes and Straughan, found in two experiments that *Ss* differed in their final performance as a function of an initial experience with different stimulus probabilities. Differences in terminal performance were also found by Goodnow and Pettigrew (1955) in a "gambling" task involving several shifts in probabilities of reinforcement during initial periods of training. The Goodnow and Pettigrew study was designed to determine whether a number of shifts in the probabilities and patterns of reinforcement would have an effect upon the strategies employed in the solution of subsequent problems.

The purpose of the present study was to investigate the effects on performance of a wide range of systematic changes in stimulus probabilities. A three-stimulus task was used in this study in order to investigate the effects of prior probability of reinforcement and to determine some of the characteristics of the variability of response found in this type of task.

Method

Subjects.—The *Ss* consisted of 225 elementary school children ranging in CA from 7.5 to 10.3 yr.[3] All *Ss* within these age levels attending the schools visited were tested.

Apparatus.—The apparatus was the same as that used in Exp. I and II.

Procedure.—The general procedure was the same as that used in Exp. I and II and reinforcements were the same as those used in Exp. I.

The procedure differed from that in the previous experiments in that different reinforcement probabilities were used. During the first 40 trials *Ss* received either 0%, 33%, 66%, or 100% reinforcement of the correct response. During the last 40 trials *Ss* either continued to receive the percentage of reinforcement obtained during the first 40 trials, or were switched to a different percentage of reinforcement. The two sets of 40 trials were presented in an uninterrupted sequence. The same knob was correct for a particular *S* throughout the experiment; only the percentage of choices on which this knob was reinforced was changed. This procedure resulted in a 4×4 design containing all combinations of initial and final percentages of reinforcement. For example, *Ss* receiving 33% reinforcement during the first 40 trials could receive either 0%, 33%, 66%, or 100% reinforcement during the last 40 trials, depending upon the condition to which they were assigned. Fifteen *Ss* were assigned at random to all but one cell of this design. One cell, 0% to 0%, was not filled. It was felt to be impractical to test children for 80 trials during which no reinforcement was received. Previous experience indicated that it is difficult to maintain children's motivation during protracted periods of 0% reinforcement. Among *Ss* receiving 33%, or 66% reinforcement, the trials on which a choice of the correct knob was reinforced were determined by 10 random sequences of 40 trials. These sequences were randomized among *Ss* tested in these conditions.

Results

When a contingent procedure of reinforcement is used in conditions involving other than 0% and 100% reinforcement it is impossible to be assured that *Ss* will receive exactly the designated percentage of reinforcement of correct responses. In the present experiment, for example, the proportions of correct responses

[3] See Footnote 2.

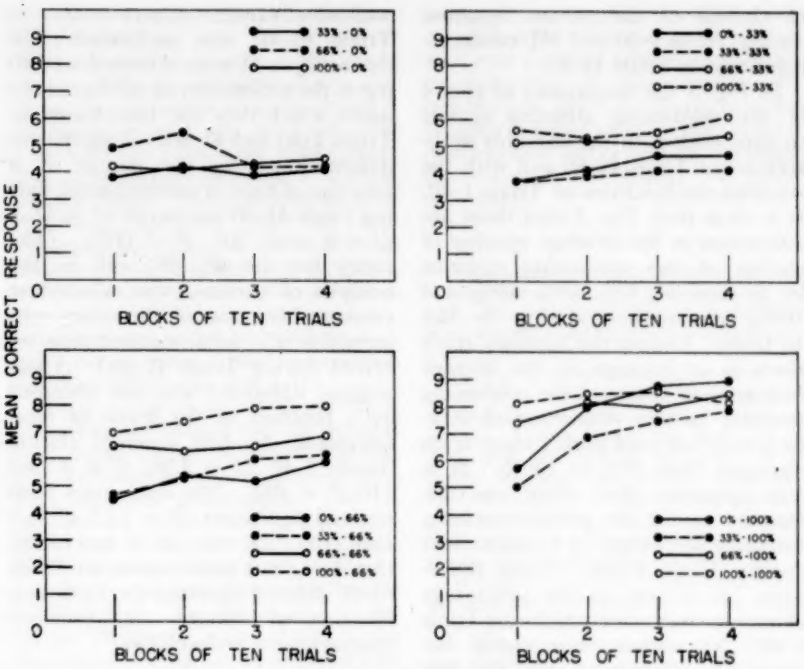

FIG. 3. Mean number of correct responses in blocks of 10 trials during Trials 41–80: Exp. III.

on which *Ss* in the 33% and 66% conditions were reinforced were .35 and .64, respectively. The designation of groups as 33% or 66% does not, therefore, represent exactly the proportion of trials on which reinforcement was received.

During the first 40 trials the average numbers of choices of the reinforcing stimulus made under conditions of 0%, 33%, 66%, and 100% reinforcement differed from each other consistently. An analysis of variance of these data reveals a highly significant difference among the groups ($F = 92.8$, $df = 3$ and 221, $P < .001$).

By the end of the first 40 trials, *Ss* in each group appeared to be approaching an asymptotic level of response. The question of whether an asymptote was reached may be answered by determining whether there was a significant change in performance between Trials 21–40 and 61–80 by groups for whom the probability of reinforcement remained constant throughout the experiment. The differences between Trials 21–40 and Trials 61–80 in the average numbers of choices of the reinforcing stimulus are not significant at the .05 level ($F = 3.83$, $df = 1$ and 42).

The results for Trials 41–80 are presented in blocks of 10 trials in Fig. 3. In this figure *Ss* are classified according to the percentage of reinforcement received on Trials 1–40 and 41–80. For example, the notation, 33%–0%, indicates that this group received 33% reinforcement

of choices of the correct response during Trials 1–40 and 0% reinforcement during Trials 41–80.

In Fig. 3 the frequencies of choice of the reinforcing stimulus appear to vary both with the stimulus probabilities of Trials 41–80 and with the stimulus probabilities of Trials 1–40. It is clear from Fig. 3 that there are differences in the average number of choices of the reinforcing stimulus by *Ss* receiving 0%, 33%, 66%, and 100% reinforcement during the last 40 trials. During the terminal trials there is an increase in the average frequency of choice of the reinforcing stimulus as the proportion of reinforcement received during these trials increases from 0% to 100%. It is also apparent that there are differences among the groups receiving the same percentage of reinforcement during Trials 41–80. These differences are related to the percentage of reinforcement received during Trials 1–40. No simple relationship between the performance on the last 40 and the first 40 trials is apparent. There was no consistent tendency for *Ss* changed from high to low or from low to high probabilities of reinforcement to fall above or below the level of response shown by *Ss* who had consistently been reinforced at the level to which *Ss* were shifted. There is some tendency for the groups receiving 33% and 66% reinforcement during the last 40 trials to be ordered according to the percentage of reinforcement received during the first 40 trials. In the groups receiving 0% and 100% reinforcement during the last 40 trials there is no tendency for the groups to be ordered in any fashion, and terminal levels of response in all groups are quite similar.

To determine whether the differences are statistically significant, an analysis of variance of performance on Trials 61–80 was performed. For this analysis *Ss* were classified according to the probability of reinforcement under which they had been tested on Trials 1–40 and 41–80. A significant difference among the groups as a function of level of reinforcement during Trials 41–80 was found ($F = 57.2$, $df = 3$ and 210, $P < .001$). (The entry for the 0%–0% cell in the analysis of variance was selected at random from one of the other cells in which 0% reinforcement was received during Trials 41–80.) A significant difference was also obtained as a function of the levels of reinforcement *Ss* had received during Trials 1–40 ($F = 3.96$, $df = 3$ and 210, $P < .01$). The interaction term was not significant ($F = 1.17$, $df = 9$ and 210, $P > .05$). It is concluded, therefore, that performance on Trials 61–80 differed significantly, both as a function of current and previous reinforcement probabilities.

A question of recent interest is the relative frequency with which *Ss* choose nonreinforced stimuli in a probability learning task (Anderson & Grant, 1957). An analysis of the distributions of choices by *Ss* in the present study was made possible by the use of a three-stimulus task. The data from all conditions were combined, since the stimulus which was reinforced was randomized among all conditions. When the left knob was correct, the middle knob was chosen on an average of 20.7 trials and the right knob on an average of 18.5 trials. The difference in the average frequency with which the two knobs were chosen is significant at beyond the .001 level ($t = 4.64$, $df = 85$). When the middle knob was correct, the left knob was chosen on an average of 14.7 trials, and the right knob on 16.0 trials ($t = 3.59$, $df = 72$, $P < .001$). When the right knob was correct, the left knob was chosen on an average of 16.7 trials, and the middle knob on 20.8 trials ($t = 6.96$,

$df = 65$, $P < .001$). The *Ss*, therefore, did not distribute their choices of nonreinforced stimuli equally. The tendency to prefer the middle knob was also found when the initial response of each *S* in this experiment was examined. On Trial 1, 56% of *Ss* chose the middle knob, 23% chose the right knob, and 21% chose the left knob. These tendencies were not found, however, when the total frequencies of choice of the reinforcing stimulus were considered. Individual t tests comparing the proportion of times the correct stimulus was chosen under conditions where the left, middle, and right stimuli were correct indicate no significant differences ($t \leq 1.41$).

These results are consistent with those of experiments showing that the terminal level of response in a probability learning task is affected by earlier experience with other stimulus probabilities. The results do not support the conclusion of Estes and Straughan that asymptotic level of response is independent of prior experience with different probabilities of reinforcement. The significance of the present finding for Estes' theory is not clear, for Estes has not made predictions concerning the three-choice case with contingent reinforcement of only one response. The problem facing a statistical learning theory in making such a prediction is that of determining some basis of distributing the effects of a nonreinforced trial between two incorrect stimuli. The present study provides evidence related to this problem: *Ss* did not distribute their choices equally between the two nonreinforcing stimuli. The results indicate a hierarchy of response tendencies such that one of the nonreinforcing knobs was more likely to be chosen than another.

Even though the results indicate that previous experience does affect final level of response, their similarity to those found in other probability learning tasks is striking, in that the level of response was dependent on the probability of reinforcement provided by the reinforcing stimulus and in that an asymptotic level of response was rapidly reached. The high dependence of the response level upon probability of reinforcement was seen in the rapid changes in performance that occurred following the change in probability of reinforcement. The rapidity with which the change in level of response occurred following the change in reinforcement indicates that *Ss* readily discriminated the change in reward probability.

Summary

Three experiments are reported concerning the effects of differences in CA, incentive, and prior experience with different probabilities of reinforcement on children's performance in a probability task.

In Exp. I, *Ss* at CA's 3, 5, 7, and 9 were tested in a three-stimulus situation with reinforcement of only one of the three responses. Reinforcement was delivered on 33%, 66%, or 100% of *Ss'* choices of the correct stimulus. An inverse relationship between frequency of choice of the reinforced stimulus and CA was found in the 33% and 66% groups. No differences appeared among the age groups with 100% reinforcement.

In Exp. II, *Ss* at CA 5 were tested under the same conditions employed in Exp. I with incentives of high or low value. In the 66% and 33% groups the frequency of choice of the reinforcing stimulus was significantly lower for the high incentive than for the low incentive groups. With 100% reinforcement no difference between the high and low incentive groups in the frequency of choice of the reinforcing stimulus was obtained.

In Exp. III, elementary school children were tested in a three-stimulus situation with reinforcement of only one of the three responses. The *Ss* received either 0%, 33%, 66%, or 100% reinforcement of the correct response for 40 trials. During a second series of 40 trials *Ss* either continued under the same reinforcement schedule, or were switched to one of the other levels of reinforcement. All combinations of initial and terminal reinforcement except the 0%–0% condition were used. Performance on the first 40 trials differed significantly among the groups and the groups were ordered according to the percentage of reinforcement received. The same ordering of the groups according to current percentage of reinforcement was found during the last 40 trials and again a significant difference in incidence of choice of the correct stimulus was found. During the last 40 trials a significant difference among the groups was also found when the groups were classified

according to the percentage of reinforcement received during Trials 1–40. Percentage of reinforcement received during Trials 1–40 did not, however, affect performance in all groups on Trials 41–80 in any consistent fashion. There was a tendency for the performance of groups receiving 33% and 66% reinforcement during the last 40 trials to be ordered according to the percentage of reinforcement received during the first 40 trials. No such tendency was found for the groups receiving 0% and 100% reinforcement during the last 40 trials.

References

ANDERSON, N. H., & GRANT, D. A. A test of a statistical learning theory model for two-choice behavior with double stimulus events. *J. exp. Psychol.*, 1957, **54**, 305–317.

ESTES, W. K., & STRAUGHAN, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.*, 1954, **47**, 225–234.

GOODNOW, J. J. Determinants of choice distribution in two-choice situations. *Amer. J. Psychol.*, 1955, **68**, 106–116.

GOODNOW, J. J., & PETTIGREW, T. F. Effect of prior patterns of experience upon strategies and learning situations. *J. exp. Psychol.*, 1955, **49**, 381–389.

MESSICK, S. J., & SOLLEY, C. M. Probability learning in children: Some exploratory studies. *J. genet. Psychol.*, 1957, **90**, 23–32.

PARDUCCI, A. Alternative measures for the discrimination of shift in reinforcement-ratio. *Amer. J. Psychol.*, 1957, **70**, 194–202.

STEVENSON, H. W., & ZIGLER, E. F. Probability learning in children. *J. exp. Psychol.*, 1958, **56**, 185–192.

(Received July 7, 1958)

Journal of Experimental Psychology
Vol. 57, No. 6, 1959

SOME FACTORS IN PROBABILITY MATCHING

IRVIN RUBINSTEIN

Walter Reed Army Institute of Research

One of the results to emerge from two-choice learning situations has been the tendency of *S*s to choose in such a way as to make their percentage of choices of an item match the percentage of occurrence of that item (cf. Hake, 1955). This has been called probability matching. The situation presented to *S* might be a series of cards—some blue, some yellow—arranged in random order with the constraint that blue occurs 75% of the time, while yellow occurs 25% of the time. The *S* is expected to predict the color of the cards as they are turned up one by one. The *S*, after a reasonably long series of trials, will begin and continue to predict blue 75% of the time, perhaps a bit higher or lower.

It has been argued that probability matching is strange behavior, and that *S* would achieve a best solution to the problem if he picked the more frequent event 100% of the time. Flood (1954) has pointed out, however, that the optimal solution assumes that *S* becomes aware that the events form a random series, but few experiments make explicit the randomness and the impossibility of a complete solution. As an alternative explanation for the optimal solution not appearing, it has been argued that *S*s place an inordinate value on correct prediction of the infrequent event (Goodnow, 1955; Grant, Hake, & Hornseth, 1951).

The present experiment is a test of these two possibilities. For the effect of awareness of randomness, the series of events is generated before *S*, so that each event can be seen as clearly independent. For the effect of the infrequent event, *S* is presented with one frequent event and three infrequent events so that there is less chance of being correct in predicting an infrequent event.

METHOD

Design and material.—For three groups of *S*s, the more frequent event occurred 67% of the time. In all three groups, *S*s were asked to predict on each of a series of 126 trials which of the alternative events would occur, and were instructed to aim at getting as many correct predictions as possible.

Group A, essentially a control group, was given a conventional two-choice task. The material consisted of 126 cards—84 blue and 42 yellow. The cards were randomly ordered on the length of runs; i.e., the run-lengths expected on a chance basis were determined and these run-lengths then randomized for order.

For Group B, the same stimulus series was used, except that 14 yellow, 14 orange, and 14 white cards were used instead of the 42 yellow.

For Group K, three playing cards were used, two Kings of Hearts and one Jack of Spades, out of a regular deck of playing cards. These cards were shuffled before each trial, and one of the three cards was chosen by a member of the group of *S*s.

Procedure.—All *S*s were tested in the same room, in groups of approximately 12 to 15 men. Responses were written on pads with numbered pages, one response to a page, and performance was monitored by either one or two assistants.

For Groups A and B the stack of cards was on *E*'s desk, but hidden from the view of the *S*s. The *S*s were told the colors that could occur, and were instructed to write their first prediction in the middle of the first page. When this was done, *E* raised the card so that it could be seen by the *S*s and at the same time called out its color. The *S*s then immediately turned the page and made the next prediction. This procedure was continued in a regular rhythmic manner taking approximately 4 sec. between presentations.

In Group K, *S*s were shown the cards that

TABLE 1

MEDIANS AND MEANS, IN PERCENTAGES, OF THE SELECTION OF THE 67% STIMULUS

Group	Trials					
	1–42		43–84		85–126	
	Mean	Median	Mean	Median	Mean	Median
A	50.48	51.19	68.88	69.05	65.00	64.29
B	44.50	47.62	66.72	66.67	69.74	71.43
K	77.61	80.95	75.80	73.81	78.12	78.57

could occur (two Kings of Hearts and one Jack of Spades) in each of the two sets of cards that were used. One *S* was used as an aide in order to impress *S*s with the fact that cards would be picked on a chance basis. Two *E*s and the aide took part in the presentation of cards. All three were seated at the desk in front of the room. The E_1 shuffled one set of the cards, handed them to E_2, and she (E_1) shuffled the alternate set. The E_2 laid the cards down separately on the desk, and the aide chose one by pointing at it. Although *S*s knew what was going on and actually observed all of the steps, they were unable to see the cards as they lay on the desk. After the aide chose a card, *S*s (who had the same type pads as the other groups) were asked to make their choices. The E_2 then picked up the chosen card and showed it to *S*s, at the same time calling out what it was. Then he took the newly shuffled set from E_1 and put those cards on the table while E_1 shuffled the set used in the first trial. (This continued for 126 trials.)

Subjects.—All *S*s were drawn from enlisted military men at the Walter Reed Army Medical Center: 44 *S*s for Group A, 41 *S*s for Group B, and 37 *S*s for Group K.

RESULTS

For the analysis the 126 trials were split into thirds giving Trials 1–42, 43–84, and 85–126 as separate units. Table 1 gives the median percentages of the selection by the three groups of the 67% stimulus. In A and B that would be blue, while in K it would be "King." Group K is seen to go well above a matching solution from the very beginning, starting out with 80.95% and ending up with 78.57%. In all cases it is well above the other two groups. Group B starts off quite low, but ends up somewhat above a matching solution. Group A follows the expected pattern of events starting at 51.19% and ending up near a matching solution.

Median tests (Mood, 1950) were done and the resulting χ^2's and P levels are shown in Table 2. Group K is significantly different from the other two groups at almost all points. In particular, when Group K is compared to Group A in the last third, $P < .01$.

The median test shows no significant differences between Groups A and B. However, over the last 42 trials the distributions (Fig. 1) of Groups A and B seem to be different with B having the greater range. In Group B, four *S*s have selected the 67% event 40 to 42 times out of a possible 42, while one has selected it only once or twice. Group A has no *S*s at these extremes. In general, as compared to Group B, Group A is more homogeneous, showing a stronger tendency to cluster normally about the mean. Some further indication that Group B's distribution does differ from a theoretical distribution more markedly than Group A's is seen in that when Group B's distribution over the last 42 trials is compared to the binomial theo-

TABLE 2

χ^2's AND P LEVELS FOR THE THREE GROUPS (MEDIAN TEST)

Groups	Trials		
	1–42	43–84	85–126
K and A	$\chi^2 = 40.101$ $P < .001$	$\chi^2 = 4.450$ $P < .05$	$\chi^2 = 10.289$ $P < .01$
K and B	$\chi^2 = 34.644$ $P < .001$	$\chi^2 = 2.601$ $P < .20$	$\chi^2 = 4.271$ $P < .05$
A and B	$\chi^2 = 2.597$ $P < .20$	$\chi^2 = 0.104$ $P < .80$	$\chi^2 = 2.002$ $P < .20$

retical distribution with 6 *df* the χ^2 is 45.434, giving a P value less than .000001, while the χ^2 of Group A's when compared to the theoretical is 11.869 giving a P value between .06 and .08.

DISCUSSION

It has been pointed out (Edwards, 1956) that when the concept of "matching solution" is applied to a group it does not account for many individual behaviors within the group. It is apparent from Fig. 1 that the individual differences within Group B are more difficult to account for in terms of the one concept of "matching" than are those of Group A. The greater range and bimodality of Group B's distribution suggest this. The factors causing it are not obvious. Perhaps the use of four colors in Group B as compared to two in Group A makes for a more complex situation where more theoretical solutions are possible to the subject. On the one hand, there is some clustering near the optimal solution which might have been caused by a "theory" of excessive failure on the *not* blue side. Secondly, there is some clustering about a matching solution. And thirdly, two *S*s are very low, as though they were waiting for the *not* blue colors to make an appearance, operating perhaps under what has been known as the "gambler's fallacy." It may well be, then, that while the median test showed

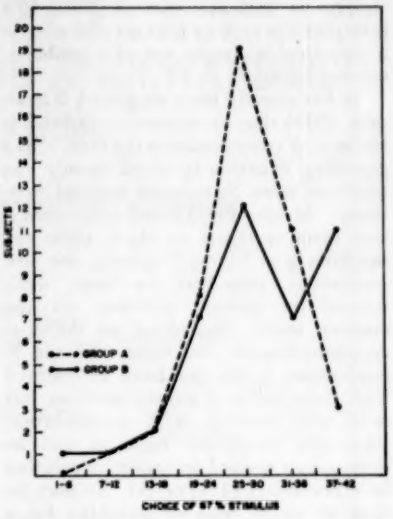

FIG. 1. Distribution of *S*s over the last 42 trials according to the number of the 67% stimulus choices.

no differences between Groups A and B, that the effect of B was more to strengthen any particular stratagem that *S* might be inclined to use.

It is clear that marked differences may be obtained from *S*s responding to a randomly generated series of the two-choice type depending upon the manner in which the series is generated. If the series is generated before *S*, so that it is clear from the beginning that the stimuli are to appear randomly, *S* does not seek a matching solution but goes well above it. Furthermore, no lengthy series of trials is needed since he reacts this way from the very beginning. On the other hand, since the optimum is rarely achieved (only three out of 37 *S*s chose "King" for the whole 126 trials) *S* must still be using other devices. At times after the experiment *S* will ask if this was an experiment in mental telepathy. This belief occurs often in experiments of the type discussed in this paper and would certainly lower the chance of an optimal solution to K. We do not wish to exhaust the factors involved, but

simply to indicate that *S* will often interpret the task to his own choosing in a way that is simply not of a problem-solving nature.

It has already been suggested (Goodnow, 1955) that an important variable is the way *S* conceptualizes the task. In a gambling situation in which money was involved more *S*s reached optimal solutions. Morin (1955) found that when *S* was given a signal to show when the machine was "lying" (giving the low percentage stimulus) he went more toward an optimal solution. In the current study the effect of different conceptualization is clear. In the K experiment *S* did not have to contend with the solution of a serial problem, but only with dealing with probabilities. Does this throw any light on why he does not go toward an optimum solution in experiments of Type A? It may be that he never tires of searching for a solution where there is the faintest possibility that one exists. What this means is that the random effects that the experimenter creates does not reach the subject. "Randomness" as a solution to a problem may not be acceptable to a subject who is accustomed to believing that if he is asked to solve a problem, the problem is solvable. Such a limitation would seem to be cultural, and within it, a matching solution is a good one.

Summary

Three experiments were performed of a two-choice nature. In Group A two differently colored cards were presented randomly with the restraint that blue occurred 67% of the time while yellow occurred 33% of the time. Group B used the same series as Group A, but where only yellow was used with Group A, three different color cards were used. In a third Group, K, a series of 126 stimuli (as in the others) was generated before *S* by the use of three playing cards. The main results were as follows: (*a*) The *S* chose the 67% stimulus a significantly greater number of times in Group K than in Groups A or B. (*b*) While Group B did not differ significantly from Group A in the number of choices of the 67% stimulus, comparison of their distributions suggests differential effects.

References

Edwards, W. Reward probability, amount, and information as determiners of sequential two-alternative decisions. *J. exp. Psychol.*, 1956, **52**, 177–188.

Flood, M. M. Environmental non-stationarity in a sequential decision-making experiment. In R. M. Thrall, C. H. Coombs, and R. L. Davis (Eds.), *Decision processes*. New York: Wiley, 1954. Pp. 287–299.

Goodnow, J. J. Determinants of choice-distribution in two-choice situations. *Amer. J. Psychol.*, 1955, **68**, 106–116.

Grant, D. A., Hake, H. W., & Hornseth, J. P. Acquisition and extinction of a verbal conditioned response with differing percentages of reinforcement. *J. exp. Psychol.*, 1951, **42**, 1–5.

Hake, H. W. The perception of frequency of occurrence and the development of "expectancy" in human experimental subjects. In H. Quastler (Ed.), *Information theory in psychology*. Glencoe, Ill.: Free Press, 1955. Pp. 257–274.

Mood, A. M. *Introduction to the theory of statistics*. New York: McGraw-Hill, 1950. Pp. 394–395.

Morin, R. E. Factors influencing rate and extent of learning in the presence of misinformative feedback. *J. exp. Psychol.*, 1955, **49**, 323–351.

(Received July 7, 1958)

ZEITSCHRIFT FÜR EXPERIMENTELLE UND ANGEWANDTE PSYCHOLOGIE

Journal of the German Psychological Association

Publisher in Germany

PROF. DR. PH. LERSCH, München; PROF. DR. F. SANDER, Bonn

Publisher for Austria	Publisher for Switzerland	Editorial Office
PROF. DR. H. ROHRACHER, Vienna	PROF. DR. R. MEILI, Bern	DR. C. J. HOGREFE, Göttingen

Vol. 5, No. 2

REINHARD TAUSCH. **The Nature of the Relationship Between the Therapist and the Client and the Verbal Role of the Therapist in Nondirective Psychotherapy**

WERNER FRÖHLICH. **The Aubert-Foerster Phenomenon and the Constriction of the Visual Field as a Result of a Moving Background**

HANS-JÜRGEN PFISTNER. **On the Contents of the Physiognomic and Pathognomic Expressive Phenomena**

ANNELIESE KRANTZ-GROSS and HANS-JOACHIM HAASE. **Writer's Cramp as Compared to Other Psychomotor Phenomena in Endogenous Depressives and Normals**

FRANZ HILLINGER. **Introversion and "Set" Among Siblings**

ANTON ZRZAVY. **A New Method of Validity Control in Personality Tests (Testdoppelgänger-Methode)**

JÜRGEN SCHMIDT. **First Steps in the Development of Methods for Psychological Tests of the Blind**

KLAUS LÖHN AND WERNER TRAXEL. **On an Electronic Sound Key**

All articles are provided with summaries in English and French.

The journal appears quarterly and contains 160 pages per issue.
Subscription: DM 64.—per volume.

VERLAG FÜR PSYCHOLOGIE • DR. C. J. HOGREFE • GÖTTINGEN
Brentanoweg 10, Germany

PSYCHOLOGICAL MONOGRAPHS

The monographs listed below are still in print—but for how long?*

ADELBERT AMES, JR. *Visual Perception and the Rotating Trapezoidal Window.* No. 324. *$1.00*

JANE SHOHL. *Effects of Exposure to Sound on Discrimination Performance in the Rat.* No. 329. *$1.00*

PHILIP WORCHEL. *Space Perception and Orientation in the Blind.* No. 332. *$1.00*

P. R. WENDT. *Development of an Eye Camera for Use with Motion Pictures.* No. 339. *$1.00*

MONCRIEFF SMITH and EDNA A. WILSON. *A Model of the Auditory Threshold and Its Application to the Problem of the Multiple Observer.* No. 359. *$1.50*

M. R. ROSENZWEIG and W. A. ROSENBLITH. *Responses to Auditory Stimuli at the Cochlea and at the Auditory Cortex.* No. 363. *$1.00*

M. B. JONES. *An Experimental Study of Extinction.* No. 369. *$1.00*

G. F. RABIDEAU. *Differences in Visual Acuity. Measurements Obtained with Different Types of Targets.* No. 395. *$1.00*

---

* A selection from Volumes 65–69. (Write for complete list of in-print items.)

Order *now* from: **American Psychological Association**
1333 Sixteenth Street, N. W.
Washington 6, D. C.